BIG DATA

THE BUSINESS REVOLUTION

决战大数据

驾驭未来商业的利器

车品觉◎著

大数据实践的先行者
阿里巴巴集团副总裁、数据委员会会长
车品觉 首部个人专著

浙江人民出版社

BIG DATA

THE BUSINESS REVOLUTION

决战大数据

驾驭未来商业的利器

车品觉◎著

浙江人民出版社

版权信息

本书纸版由浙江人民出版社于2014年4月出版

作者授权湛庐文化（Cheers Publishing）作中国大陆（地区）电子版发行（限简体中文）

书名：决战大数据：驾驭未来商业的利器

著者：车品觉

字数：146000

电子书定价：27.99美元

BIG DATA: THE BUSINESS REVOLUTION

BIG DATA

The Business Revolution

前　言

忘掉大数据

我们正处在历史的转折点上，数据技术在快速变革。大数据成了人们竞相议论的热词，但鲜有人提及这场巨大变革中人们需要具备的能力。无数的企业及个人望“数”兴叹：“大数据与我何干？”未来是大数据的时代，未来的竞争就是数据的竞争。也许，我们早该忘掉那些华而不实的喧嚣，让大数据真正从“看” 到“用”，真正“活”起来。

大数据的力量来自触类旁通的关联。我们以前总是用数据来证明或企图说服工作上的盲点，而如今的数据不再成为一加一的依据，而是具备了预测和开创新机的能力。

用数据找机会

我第一次见证大数据的魅力是在15年前，只是当时的我并不知道那就是“大数据”。20世纪90年代初，我认识了一些以博彩为生的朋友，这些人组成了一个团队，每年通过赛马，就能盈利数亿港元。我非常惊讶，要知道很多人在赌马场上可是血本无归的，而他们却能把这种概率游戏变成稳定的盈利工具。原来，他们的秘密就是使用了一套“养数据”　策略——将每一场赛马比赛的过程都录了下来。当时我觉得这个做法很奇怪：“电视上已经在播放录像了啊，还另外录比赛干什么？”后

来我得知，他们居然在每场比赛中都会录取赛马不同角度的录像。通过这些录像，他们分析出骑师、马匹有哪些失误动作，这些动作会带来怎样的后果，然后再把这些数据“清洗”出一个更准确的数据（Smart Data）。赛马过程中有许多意外，他们利用数据来还原——如果在没有意外发生的情况下，马匹在不同场地与不同骑师配合中的应有速度。就这样，他们可以更准确地判断出每匹马的实力和获胜的机会；就这样，通过悄无声息的数据收集，每年入账数亿港元。

令我最为震惊的是，他们竟然不看表面数据，而是从无限数据的机会中寻找核心数据。

这正体现了大数据与以前数据最大的不同。以前，我们都是有问题找数据，而大数据时代，其最核心的特质则是“用数据找机会”。我们做大数据，必须要有一个预判，那就是哪些数据是你必须要提炼出来来解决盲点的。赛马的结果其实充满了“意外”，新的数据角度帮助我们一窥真实的结果，这就是“用数据来还原真实”。

只有实效的数据才是正道

现在，大数据的概念纷繁复杂，媒体上充斥了各种关于大数据的报道，但其中不乏牵强附会、滥竽充数的言论，某些媒体甚至把简单的统计也冠上了“大数据”的头衔。

作为一个跟数据打了十几年交道的人，我深深地知道从“看”到“用”，再从“用”到“养”的运营数据，本身就是一个复杂的过程，而也许目前我们最应该做的，就是暂且忘记大数据的概念。行胜于言，只有具备实效性的数据才是正道。我希望从一个实用的角度来拨开大数据的“迷雾”，告诉每个人大数据的具体运作应该是什么样的；而且数据量绝对不是一个最重要的问题，我们要的不是数据的量，而是有“质”的量，这正是我

写作本书的重要目的。

数据，决胜未来的商业利器

在这个风云变幻的数据时代，只有让数据成为商业的利器才能决胜千里。

首先，我们需要拥有一套具有商业敏感度的数据决策框架，可以使企业“看”得更准，并能够对近期做了什么是对的、什么是错的进行判断。这样一来，快速的数据反馈可以让每个决策的误差得到适时修正。其次，让数据真正从“看”到“用”，让用数据成为构建企业生产力的重要部分。再次，让Data Technology（DT战略）深入到企业的每个角落，使数据从生产、收集、使用、分享到反馈变得简单易用。最后，让DT战略落地还要特别注意数据的稳定、准确、时效和有效实施。

在阿里巴巴，我学习到一个很重要的经验——人和事是分不开的。企业要想成为一家数据化的公司，文化的培养必不可少。“混、通、晒”及“存、管、用” 两套内功是让企业的血液（数据）流动起来的关键。数据流动得宜，则神清气爽；相反，数据如果出现停滞或质量问题，企业则性命不保。

锻造数据力和思考力的合力

大数据时代的到来，让我们几千年来第一次碰上了数据化带来的机遇和滚滚红利，也让每个人的思维方式出现了重大改变。很多时候，我们欠缺的不是解决问题的方式，而是定位问题的能力；我们应该思考，有没有数据可以改善我们的盲点；我们该如何学会用“假定数据是可获取的”来重新思考周围的一切。当这种数据化思考成为你的“直觉”时，就能够把数据的力量和思考的力量融合在一起，产生出新的无与伦比的合力了。到那时你会发现，周围的一切都将因此而焕发出新的生机。

最后，衷心感谢一年来帮助我完成这个看似不可能完成的任务、写完这本看似不可能完成的书，最终付诸实践的老师及朋友们：

感恩给予我写这本书勇气的雪谦寺拥珠扎西仁波切；

感谢一直帮助我推动这本书前期工作的林君君；

感谢用业余时间协助我完成写作及案例工作的赵军、李江、吴慧敏、欧吉良、张大红；

感谢湛庐文化专业的策划和编辑团队；

更感激家人一直容忍着我的忽略……

正如数据是一种信仰，万法唯心，每一刻的进步都来自真诚的对待！

品觉

于西子湖畔

2014年4月

特别说明

1. 本书观点乃作者十几年经验心得的分享，只代表个人立场。

2. 本书的缘起来自解决雪谦寺重建的经费及桑珠助学的资金问题，本人的所有收益将全部捐出作以上用途。

目录

第一部分

从数据化运营到运营数据

数据十诫

1 好的问题，答案就在里面。

2 在实践中提炼数据。

3 让数据（Data）变成科技（Technology），惠及更多人。

4 让数据跟着“人”走。

5 木有数据质量，神马数据都是浮云。

6 以“假定数据是可获取的”去思考问题。

7 大数据安全，不是监管。

8 利用数据拿到更有用的数据。

9 建立数据的数据，才有进步。

10 让人做人擅长做的事，让机器做机器擅长做的

事。

01 大数据，为什么很多人只会谈，不会做

BIG DATA 众说《决战大数据》
The Business Revolution

林奕彰 eBay 大中华区CEO

我跟品觉相识多年。在大数据还没流行以前，他就开始关注这个趋势，研究数据应用的意义以及怎样真正读懂数据和利用数据。这本书深入浅出，把深奥的大数据讲得生动有趣，非常好读；书里面集中了很多品觉个人的真知灼见，非常值得一读。读此书是一种享受，正如我享受与品觉的友谊一样。

有人说，大数据就像国王的新衣，每个人都在国王面前说着动听的话，国王信以为真，其实竟然不知道自己在裸奔。

的确，网络上有很多人在谈大数据，但是他们只会谈，不会做，因为他们根本就没有做过，包括那些所谓的“大数据专家”，他们真的做过吗？没有。

事实上，这些人对大数据内在的问题一点儿都不了解，更别说知道大数据的水有多深了。

目前，在大数据方面，无法深入应用的原因在于，从收集到使用的大数据价值链出现了问题。从理论上来说，从收到用的螺旋式循环是一个巨大的涡轮，只有先数据化运营，然后才能运营数据。而现在的情况是，用数据的人不知道大数据从

哪里来，做数据的人不知道大数据如何使用。用的人不敢用，因为大数据的真实性；做的人不知道怎么用，因为大数据的复杂性。这一问题造成的结果就是，数据量变得越来越大，而且越来越无法有效地使用。

大数据从来不是免费的午餐

> 大数据从来不是免费的午餐。大数据的来源是多种渠道的，偏倚、随机的误差总是存在。

我先问一个数据管理上最现实的问题："大数据如何备份？"毫不夸张地说，大数据已经这么庞大了，如果再备份一次，你的成本起码会增加一倍。

做大数据基本上都要从大量地收集数据开始，因为这些数据在未来会大有用处。但是，你是不可能无止境地收集下去的。在这里，你已经看到了一个再清晰不过的伪命题——大数据的确能够备份，但是成本会增加两三倍……然而，"以前重要的数据肯定都需要备份啊，"你自然会问，"如果不能备份，我该怎么办？"而这就是一个大数据管理中必然要遇到的问题。

我再问一个问题，如果你在数据使用方面一直得心应手，整个商业链条和数据紧密相关、相辅相成。但是，现在数据链忽然断了，或者不再有效了，你该怎么办？

你需要注意的是，这一问题说的不是你有没有使用好数据，而是说曾经你可以得到的数据现在无从获得了；或者说这些数据不能再在线上收集，只能从线下获取，成本也就相应升高了；抑或说，这些数据存储在其他地方，你不能使用了。在诸如此类的情况下，你该怎么办？

当然，如果你对于数据的使用本来就很生疏，而且也觉得

无关紧要，那么数据链消失了、断了、失效了，你也不会有太大烦恼。但假如正是你将数据使用得风生水起的时候，如果出现这样的问题，那你也就只能束手无策，眼睁睁地看着机会溜走。

大数据从来不是免费的午餐。伴随着大数据热潮的到来，关于大数据的一些新问题层出不穷——大数据会夹杂着虚假信息；大数据的数据量很大，但有用的信息不一定多，甚至还会破坏核心信息；大数据的来源是多种渠道的，偏倚、随机的误差总是存在。

但是，我们也需要客观地认识到，大数据现在面临的这些问题，其实就是把小数据中的一些问题放大了。小数据中难道就没有噪音会破坏我们的核心信息吗？当然也有，只不过当大数据把数据量放大和变多的时候，噪音的破坏性也相应增大。小数据中难道就没有渠道偏倚和随机的问题吗？当然也有，但是在大数据的背景下，问题就被更明显地放大了。

人的断层

断层才是大数据所面临的最严重的问题。现在，收集数据的人并不清楚未来使用数据的人要做什么，这是目前大数据的一大关键命门。

说了这么多，事实上还没有触碰到如今大数据面临的最大问题：人。

很多人都会问，大数据能带来什么价值？怎么衡量大数据创造的价值？事实上，最直接的衡量标准就是，在经营上它为你赚了多少钱，带来了多少实际的利润提升。

对于这一问题的解决，现在很多人最倾向使用的方法是计算“在用了大数据之后，点击率提高了多少，转化率提高了多少”。但是要知道，转化率和点击率能提高的数据，可能根本

不是投资人或者公司最高管理者对大数据的期望。对于业务人员来说，转化率能提升5个百分点就已经非常好了，如果将转化率从2%提高到3%，简直就是奇迹了，但对于公司最高管理者来说，也许这并不是他想要的大数据。

你需要认识到，断层才是大数据所面临的最严重的问题。现在，收集数据的人并不清楚未来使用数据的人要做什么，这是目前大数据的一大关键命门。

在使用大数据时，我们通常的做法是先把数据收集起来，因为我们知道在未来的某一时刻，这些数据对我们可能有用。不过，“未来可能有用”就注定会引发一个问题——收集数据的人不知道未来使用数据的人要做什么。这时候，如果你再问收集数据的人“如何才能更好地收集数据”，那么，数据的使用就会陷入一个死循环。

事实上，不仅是收集数据的人，就连使用数据建模的人，同样也不清楚当前的数据是如何获得的。数据建模是数据使用的关键环节，使用数据建模，就是根据以往的经验，从中寻找到一些潜在的规则，然后把这些规则结合起来去解决问题。

> 举个例子来说，我现在身上只有10元钱，我可以用它买一本杂志，也可以用它坐几次公交车，或者可以用它来买方便面充饥，这三种选择在一般情况下都是可行的。但是，如果再加上一个“我没有吃早餐”的场景时，在以上三种可能性中，我选择去吃方便面的概率自然会比较大。这其实就是一个简单的模型——输入“拥有10元钱”和“没有吃早餐，肚子饿”这两个场景，输出“买方便面”这个结论，它帮助我们快速地选择了一条解决之路。

选择“吃方便面”的这个场景虽然看上去很简单，但同样是一个经验的总结，这个经验就是“肚子饿了要吃饭”。把建模这个过程说得更加复杂或者专业一点，就是基于很多以往的经验，进行总结，或者是对旧有数据使用规则的发现，将经验和

数据相结合，最后输出一个可以被套用的业务规则。

不过，此时使用数据建模的人并不一定完全清楚数据是如何产生的，这是一个客观存在的难题。因为就大数据研究来讲，我们很多时候使用的都是他人的数据，而他人的数据你又怎么能够保证会完全清楚地知道呢？

现在，在运用大数据时流行一个例子，比如我在搜索引擎里发现某个地方搜索“感冒药”的频率非常高，于是，我就断定这个地方可能出现了流行性感冒。而这个数据源是从何而来的？我虽然知道很多人搜索了“感冒药”这一词语，但是却完全不知道在搜索引擎里搜索这个词语的人到底是谁。而提供数据的人既没有责任要告诉我数据是怎么来的，也不必告诉我数据的质量如何，更不必告诉我数据到底会不会有偏差。这样的结果推及到大数据领域就是，研究数据的人完全不清楚数据是怎么来的。

模型数据从何而来

创建模型的人也不知道自己所采用的数据在未来是否稳定，而使用模型的人也不知道整个数据的来路或加工过程。

当使用模型的人不知道模型数据从何而来时，其使用模型的角度就要发生改变，具体来说就是视场景而定。比如说，今天你要来杭州的淘宝城，有人告诉你“今天你应该从文一西路过来”。你问原因，他回答说：“因为模型是这样提供路线的。”他不会跟你解释，就说模型是这样的，因为下午2点钟到4点钟，从文一西路过来比较快。这个例子就是模型使用的一个场景，选择文一路则是模型的一个输出。那么，你要不要走其他的路？走文二路或者文三路？可是，这种选择的结果就是未知的了。

模型对于很多人来说是一个黑盒子，充满神秘性和未知

性。所以，即便是模型创建者将模型公开，也会因为创建者和使用者专业能力和知识背景的不同，使二者之间出现信息不对称的情况，使用者自然也就不知道为什么会选用此种解决方法。也就是说，创建模型的人可能不知道此种模型效果好不好，而使用模型的人也不知道该怎么去反馈使用的结果。这样一来，这种信息不对称会越来越严重。

这一问题恰恰体现了如今大数据实践中非常严重的断层问题——不只是收集数据的人不知道将来的人怎么使用数据，就连创建模型的人也不知道自己所采用的数据在未来是否稳定，而使用模型的人也不知道整个数据的来路或加工过程，这些都是普遍存在且很现实的问题。

从公司管理层的角度来看，投资人了解数据的意义是什么吗？高层管理者对数据的期望和中层管理者之间又有什么不同？他们知道数据能帮助企业做什么吗？这些问题的答案完全会因立场而异。

中层管理者大都不知道数据能帮助他们做什么，他们没有管理者的视野，相比之下，你只需要告诉他们数据能解决什么问题即可。相反，数据分析师可能就会更加困惑不解："我做了这么多东西，为什么你们不用？"

每个层级和功能部门都是一个断层，而且对数据价值的内在衡量都不一样。所以，当我们讲到数据价值时，没有人能对此给出一个合理的定位，原因就在于有几个关键问题没有区分清楚。一是要明确这是谁心里的数据价值，投资人、管理者、中层、数据分析师们心中对数据的价值自然不同；二是要明确数据的分类，不同类型的数据所产生的价值各不相同。为什么我们在大数据应用方面存在障碍，一个重要的原因在于，应用人员对于数据价值和数据分类没有明显的界定。对于这个话题的探讨，我们会在随后的章节中重点讲述。

更主动的管理，更多的创新

> 这是一个“从用数据到 养数据”的过程；这是一个“从数据化运营到运营数据”的过程；这同样也是一个“从看到真用”的过程。

在当下的大数据环境里，数据其实与商业模式密不可分，每个人都认识到它的经济价值是巨大的，但今日的大数据发展趋势之快，　对于很多公司来说，变得更加虚无缥缈，难以把控，让每个人抓狂，让每个人手足无措。

而在这一方面，阿里巴巴已经对数据化运营做了不少有益的尝试。从2005年开始，淘宝有了第一个数据分析师，并一直致力于一件事情，那就是用数据来帮助企业运营和解决问题。但在不断使用数据的同时，也发现了数据本身的问题——大数据需要更主动的管理，也需要更多的创新。

数据化运营就是用数据去解决问题，但是如果我们想把数据做得更好，解决更多新的问题，就需要去做一件以前未曾做过的新事情——运营数据。对于阿里巴巴来说，这件事情是从2011年才有计划地进行的，企业主动收集数据，并且以此去创造更优质的新数据，让新数据更好地服务于企业的运营。这是一个“从用数据到养数据”的过程；这是一个“从数据化运营到运营数据”的过程；这同样也是一个“从看到真用”的过程。

从数据化运营到运营数据是一个循环，今天的电商企业正走到了其中的一个节点上。在经历了起初大数据的喧嚣之后，大家终于感受到，要使大数据产生真正的商业价值，我们要关注的内容并非4个V[1]那么简单，而应该将焦点放在如何真正让数据落地之上，即从数据化运营到商业管理能力的提升。

[1] 4个V，是大数据的经典定义，包括海量的数据规模（Volume）、快速的数据流转和动态的数据体系（Velocity）、多样的数据类型（Variety）和巨大的数据价值（Value）。——编者注

BIG DATA
The Business Revolution
数据化思考

问题就是答案

大数据改变了人类学习、工作和生活的方式，更改变了人类的思维方式，而今天我所谈及的数据化思考正是在大数据的特色中产生的。

西方式的思考着重于系统性分析，因此对一件事情的理解过程，大都是利用数据来对之进行细分/归类、对比、溯源，以及从三者的动态趋势中去找出变化的规律。我们可以从以下几点中看出：

- P——目的，我们的目的是如何让自己更“好运”。

- I ——定义，我们要定义什么是“好运”，例如可以细分为：财运、健康或者人缘。

- M ——量化，如何量化“好运”？请参照前面的定义寻找可量化的目标：工资增长率、体检中某项指标的高低等。量化了之后，我们可以从趋势中预判出做什么可以更“好运”。

- A——评估，执行后，评估什么做对了、什么做错了。

可见，西方式思考中的溯源，就是从事物中找出因果关系，寻找能够有效地获得成功的驱动力；而东方式思考中的溯源则是从本质找出根源，例如，“好运”的本质可能在经过分析后会体现为“快乐”，但量化出来的结果却是“笑”。

然而，不管是东方，还是西方，要溯源，一切答案都得从问题开始。因此，良好的提问技巧能使我们在寻找答案的道路上事半功倍，好

问题（Deep Question），可以使我们离答案更近一步。有一位知名CEO曾很得意地问了我老师一个问题：“你觉得到底是eBay的商业模式好，还是Amazon的商业模式好？”其言外之意是，今天Amazon比eBay做得好，是不是也就意味着它的商业模式比较好。我的老师并没有直接回答这位CEO的问题，而是反问了一个问题：“你的问题也许跟商业模式无关，产生这样的结果会不会只是因为Amazon的CEO比eBay的CEO做得好呢？”无须多言，问题就是答案。所以，对一件事情的认知，要看你是否具备了用问题解决问题的能力。

2011年的时候，我所在的公司要从数据化运营开始转向运营数据。当时，我负责支付宝的运营数据工作，却很是苦恼，毫无头绪。我没有任何思路，于是我打电话求助我的老师，咨询他对于这个问题我应该如何思考。

接下来发生的事情让我非常难忘，我的老师并没有直接回答我，而是问了我三个问题。

> 现在你所在的公司，面对的3大问题是什么？公司未来3个月中，要解决的问题是什么？在过去的1个月中，你做对了什么，做错了什么？
>
> 我现在给你5分钟时间，如果你说不出来公司目前面临的3大问题，而你作为商业智能部的负责人，就基本上不该继续坐在这个位置上了。

当然，我明白这3个问题里面已经包含了答案。当你要做数据化运营或者运营数据的时候，你必须要问自己这些问题。如果你没有问自己这些问题就去想如何收集数据的话，那肯定会出问题，这是很关键的一点。所以，思维方法很重要，在有了数据、有了对商业的理解的前提下，思维方法是不可或缺的关键。思考的方法是无穷尽的，就像两个武林高手过招，假设一方面对的是功夫高手李小龙，作为局外人的我自然无法给出应该如何胜利的建议，而是需要他自己动态地去寻找答案。

寻找答案的过程中，有一个很重要的衔接点。很多时候，我们会受一些固有问题的影响，如果不懂得变通、不懂得如何提炼核心的话，就会产生思考的盲点。如果你的思考出现了盲点，就注定问不出一些非常关键的问题，就好比，对方明明是男性，还问他是不是男性，就显得

很没有价值了。

有时，有些东西并非是你认为的那样，所以，你才会连问问题都懒得思考，就问出像前面那位CEO的“eBay和Amazon谁的商业模式更好”那样奇怪的问题。

话说回来，在知道自己的思考有盲点的情况下，即便问题问得不好也并不丢人。但如果自视很高，什么问题都不问，盲点注定会相伴一生。我曾经见识过一位顶级CEO，其厉害之处就是，凡事都说：“我不懂，麻烦你再解释一下。”

最重要的是思考，尽管你懂，但如果在问问题、看问题的角度上没有拓宽视野的话，想解决问题依然很难。就好像咏春拳里的“问手”，它并不能用于过招，而是寻找答案的一种方式。你一碰，它就有答案。你不动，它不动，而且它会避开，你只能寻找破绽再出击。“问题”是用数据来拿数据，而“问手”就是用一个问题引出另一个问题。

在大数据的商业环境里，要既懂数据，又懂商业，还要拥有一套好的思维方法，而数据化思考正是这样一个崭新的事物。

当我们想知道问题的答案时，就要首先学会如何问对问题，因为一个好问题会引发出很多答案，这是一门艺术。在我看来，人的思考方式在大数据充斥的商业时代尤为重要。在后面的“数据化思考”中我将谈到一些有关思考的例子，以供大家参考。

答案不重要，思考的角度才重要。可见，要习得一套巧妙的数据化思考方式，三分靠想法，七分靠实践。所以，切勿空谈。

02 大数据的本质就是还原用户的真实需求

BIG DATA 众说《决战大数据》

The Business Revolution

沈 亚 唯品会创始人，董事长兼**CEO**

在我眼中，品觉一直是我非常尊敬的国内数据领域的大师级人物，这是他的第一本著作，也是汇聚了他十几年数据行业经验，呕心沥血的精髓之作。他一直强调数据是尊重人性的，只有有商业实效性的数据才是有质量的数据；他一直强调不能只是纯粹地看数据，要用数据还原真实的现实与场景。这些见解不仅对每个企业的未来发展具有极大的指导意义，而且对每个普通人建立数据化思维和进行个人数据管理都有很好的启发作用。这本书贵在坦诚和实用，每个人都能从书中收获颇多。

当我们身边的所有人都在兴奋地讨论着大数据时，你有没有心存不解——虽然我们每天都在收集数据，但我们为什么要关注它们？这些数据又如何能够为我们所用？

的确，当回归到“数据能够帮助你产生什么价值”这个问题时，如果你自己都没弄清楚，那么就算给你再多的数据，你也只会手足无措。我们身边的大数据现象越来越普遍，数据的广度和深度都呈现出几何级甚至指数级增长的趋势。在思考数据的价值时，我认为，可以从三个维度来考虑。

首先，你能否清楚地识别（Identify）用户的身份？其次，你能否搞清楚收集的数据对你的价值（Value）是什么？最后，收集数据时的场景（Situation）是什么？

识别，让似是而非的行为数据串联起来

企业有多大的能力去识别一个“碎片化了的个人”，将是一个巨大的考验和机会。

我先来说说识别。现在，很多人都同时拥有手机、个人PC、平板电脑等多个数码设备。比如，某个人拥有两部手机、1台iPad、3张信用卡，这个人每天都登录你的网站。在这种情况下，你能不能知道这是同一个用户？你有没有这种识别能力呢？银行用户的信用卡虽然可以识别静态个人的身份，但银行缺少当事人使用电脑、手机或平板电脑等更广泛的网上浏览设备时的身份识别能力。

这时，有谁能够识别这三个设备和三张信用卡背后的用户都是一个人？而目前，企业有多大的能力去识别一个“碎片化了的个人”，将是一个巨大的考验和机会。

在识别过程中，分辨用户其实并非难事，比如，企业可以查看用户电脑里的cookie，依此获取用户的E-mail地址、网购时的收货地址，甚至私人手机号码、亲密联系人、信用卡与身份证信息，等等。而以上这些信息都可以作为鉴别“碎片化了的个人”的依据。

需要注意的是，这些属性有些是唯一的，有些却不是，但它们依然有很大的参考价值。比如说，你经常会将手机借给别人使用吗？经常会将电脑借给别人使用吗？当然不会。那么信用卡呢？想必应该会更少外借了吧。

了解了识别身份的方式之后，我们就可以把散落于网站内外似是而非的行为数据串联起来。不同的网站，收集到的数据是不同的，数据的价值也千差万别。如果某个网站连接了你的三台设备和三张信用卡，那么这个网站收集到的内外数据就很容易将你识别出来。所以，是否知道用户是谁，决定了企业数据收集行为的意义大小。

价值，企业价值Vs客户价值

从企业价值来看，数据收集实现的是企业资源的合理分配；从客户价值来看，数据收集实现的是顾客体验的提升

数据收集的价值包含两个维度。一方面，你是否能衡量这个数据对企业产生的价值——你不会将用户的所有行为都记录下来，而是记录那些对企业自身有帮助的数据，即企业价值；另一方面，你是否能衡量这个数据对顾客的价值——这个数据如何帮助企业为客户提供更好的服务，即客户价值。

从企业价值来看，数据收集实现的是企业资源的合理分配。例如，把推荐系统做得更好，让更多的用户可以有更高的概率找到他们想要的商品。这样，就能提高商品的购买率，为企业创造更大的经济效益。

从客户价值来看，数据收集实现的是顾客体验的提升。例如，便捷的搜索引擎可以使顾客更容易地找到他们所需要的商品，为他们创造更加优质的购物体验。

数据提供的价值，从不同维度看会有不同的结果。比如说，每天下午两点钟左右，招商银行门口会有很多客户排起长龙。如果以客户价值为核心，那么思考的角度就应该是每个在下午两点钟去招商银行办业务的用户，平均到底需要多少分钟才会得到相应的服务。但是，招商银行的做法是对大量数据进行研究，目的是想得出在下午两点钟时如何减少排队的人数，这就是以企业价值为核心的考量。

以企业价值为核心去审视就是，如何帮助企业在下午两点钟时分流排队的人数；以客户价值为核心去审视则是，如何让客户减少等候的时间，并得到所需的服务。同样一件事情，这两者的提炼是不一样的：一是思考能否缩短在这家银行排队的时间，让用户提前两分钟就能享受企业的服务；二是思考能否

使用手机来解决排队问题，通过手机银行让实体银行里的排队人数减少。

我的观点是，数据价值的确要分是从谁的角度出发的。所以，当我们谈大数据价值的时候，第一点要注意的就是角色不一样，对于数据价值的看法也就不同，所以在衡量价值时要考虑到受众和给予者这两个对立面的不同看法。在招商银行的例子中，后者更多地是看，在两点钟有20多名用户在排队，而受众哪里会对排队人数产生敏感啊，他们只会觉得若10分钟内没有得到服务，就是等了很久了。

从电子商务的角度来看，推荐系统的做法跟银行类似。现在，所有的电子商务公司在考核推荐系统时，都是以点击率作为KPI[1]的。

从企业考核推荐系统的角度来说，点击率是主要指标。它们想的是如何让用户更多地点击自己的产品，而对于消费者来说，却并非如此。用户想的是："今天我有一个很想买的东西，我能不能买得到呢？"

所以，在使用数据改善用户的购物体验时，企业应该是仅仅改善推荐系统，还是要了解用户整个购买决策的链条结构？可能的答案是，企业需要知道用户在登录网站之前的情况，而不是登录之后的情况，这就需要企业全面客观地去看自己的网站是不是有能力还原用户的需求了。

场景，你知道当时所有的场景吗

场景与还原并行——前端还原消费者场景，后端还原业务需求。

在准确识别和衡量了数据价值之后，我们再来看一下数据收集时的场景。现在，大数据的价值吸引了很多企业和组织的兴趣，因此它们非常重视自身的数据收集与存储。但是，在现

实操作过程中，问题仍然层出不穷。大量的碎片化数据是噪音，让事实串联的行为本身变得非常困难。而值得思考的另一方面是，当我们把这些枯燥的数据串联起来时，就一定能代表事实吗?

数据分析实例

> 早上，你在上班的路上，看见有个男生穿了一件非常好看的T恤，你心动了。于是，你来到公司，坐在座位上的第一件事情，就是在购物网站上搜索T恤。结果呢？出现了10万个相关商品。
>
> 你正在犹豫怎么挑选的时候，老板突然在办公室里宣布，大家集体去会议室开会。你坐在会议室里，发现开会好无聊，于是打开了手机应用，继续想怎么买到那款T恤，然后你筛选了一些比较中意的品牌，但是依然没有找到那款你在路上碰到的、中意的T恤。
>
> 最后，手机上弹出了一个手机促销活动的广告，你发现一款你非常喜欢的品牌的手机价格很便宜，虽然你已经有手机了，但是你依然毫不犹豫地买下了它。

那么，在向你推荐产品时，一家公司到底有多大的能力，才能还原你所处的场景呢？

搜索引擎暴露出你目前住在上海的事实——你是最近搬过来的，网上的收货地址换到了上海。所以，它更多地推荐了上海的卖家给你。而你在到公司上班之前，才看到了那件T恤，觉得很好看，所以决定搜索T恤。但是，在你没有登录电商网站之前，T恤这个因素只有你知道，电商不知道。

因此，当你在搜索引擎里描述一件T恤的时候，你跟这家电商的第一个接触点就出现了，所以，电商还原的场景是“早上九点半，你搜索了T恤，搜出了10万个结果，但你一个都没有点击进入”。电商完全不知道，这时候你不点击的原因仅仅

是老板在叫你开会。

而会议开始后不久，你打开手机应用再次搜索T恤，这是和此电商的第二次接触。那么，后来在无数的品牌里面，你为什么忘记了早上喜欢的那款T恤，最终选择了一部比较便宜的手机？有谁知道你为什么这么选择？最后，你买了一部手机，选择了上海的卖家，事实上你只是过去在上海出差了几个月，而现在却在杭州。

试问，一家公司能有多大的能力猜到这些复杂的场景，以及这些每一天每时每刻每一个进入网站的用户背后发生的特别故事？

所以，每一个数据分析师都闭着眼睛说：“我不管用户之前看过什么T恤，反正他搜索了之后什么都没点击，可能是推荐的算法不给力。”而在手机端中，又有一名数据分析师说：“这是这个用户第一次登录无线终端啊，也是第一次点击了T恤搜索选项。”

正如你看到的，在本应有那么多连接性的数据里面，在购买T恤的案例中却没有了任何连接性，那么，企业怎么还原你的购物场景呢？每个人都在猜想可能的原因。而且，值得一提的是，这个案例仅仅涉及了用户在一个网站上做出的购买行为，且没有做过全网比价，而在真实的购物场景中哪里会这么简单？即使是这么简单的一个场景，企业又该如何实现还原呢？

每天都有大量的碎片化数据产生，每天我们都在假定。那么，用这种状态分析数据，能有多可靠？

所以说，企业在收集数据的时候，一定要明确，自己是否有能力去辨别用户，是否有能力去收集其在你的网站中发生的所有行为，是否有能力把手机端和PC端区分开来？另外，企

业又拥有多大的能力看到具体的场景是怎样的，比如说，今天南方有没有下雨，东北有没有下雪。归根结底，我想说的是，作为企业，你到底有多大的能力去还原用户的真实需求？

还原是一个瞄准器

数据的本质就是还原，这是收集元数据的关键方法。

当我们在进行用户的场景还原时，必须认清数据收集的领域是什么。在不同的领域里收集到的数据，可以找到与其所在领域里不同的东西，比如，搜索引擎和社交网络（SNS）得到的数据就是不一样的。而企业首先要做的是，确认用户是不是同一个人，比如在SNS里涉及的很多信息主要都是聊天内容，如果我是做SNS的，我就会更多地去寻找这个人和其他人的关系。他今天跟张三聊了3分钟，明天跟李四聊了5分钟，这项数据在SNS领域里可以获得。但当我们要真实地还原整个人的行为的话，最好要有不同领域作为互补，这会让你掌握更多更全面的信息。

你有多大的能力知道哪些数据是同一个用户的，这是企业首先必须解决的问题。然后再去关注，收集到的数据的量这么大，广度这么宽，价值在哪里。而当企业不清楚收集到的数据是不是同一个用户的时，那这个数据又有什么用？所以，在大数据里，最重要的还是收集人的数据。

而数据的价值，正如我们前面所分析的，必须来自场景。

对于消费者数据的收集，其中一个瞄准器就是你能否还原用户购买行为的场景。基于科技的不断进步，如果有一天Google 眼镜成为每个人的标配，或者是有一天，我们买的每一部电脑，其本身都是跟手机捆绑的，那么这两种交叉数据是很容易获取的。

为什么场景会变得如此重要？场景是不是被准确地表达了？场景是否会成为一件事情的背景，用来还原整件事情？有一年的“十一黄金周”，我们发现很多用户使用iPad购物，为什么？你或许不会想到，这是因为那年的“十一黄金周”第一次实行黄金周高速公路不收过路费的政策，很多人堵在了路上，没有其他事情做，所以只能使用iPad购物。如果企业在分析数据的时候，没有考虑到10月1日整个中国的高速公路都出现拥堵的这个场景，企业就没有办法还原整个场景，也就无法解释这个现象。

我还发现了一个新场景。有一天，我们研究了一些无线数据——用二维码让用户到达我们预想让他到达的页面。我们可以看到，iPhone手机在扫完二维码后就到达了页面，但是安卓却没有。在中国，很多装有安卓系统的手机在扫完二维码之后都无法自动跳转到关联页面。这时，安卓手机就成为用户登录网站购物的重要场景。不管是一部iPhone手机、一部三星手机，还是一部其他类型的手机，手机的大小和系统的兼容性本身都能对场景产生巨大的影响。而如果我们盲目地去观察数据本身，自以为是地认为用户没有进行点击，那就大错特错了。事实上，是他点不了。

可见，有很多看似无关紧要的东西都在场景里，而在无线移动终端的世界里，这个场景又平添了很多其他的东西，这都需要我们仔细地甄别。

当我和数据分析师们聊天时，我总是会特别提醒，在研究无线数据的时候，要特别注意的是用户在每天移动的时间点和非移动的时间点里都做了什么。从起床到睡觉，有几个时间点基本是固定的，所以基本上就可以判断每个人一天的行为走势是什么。

我曾经在一次电商大会的圆桌论坛上提出了一个观点，当

时与会者都很认同，就是我建议把一些以前用来观察用户忠诚度的框架，比如RFM模型[2]来做收集数据的瞄准器。有什么数据能让我更好地看到R，更好地看到F，更好地看到M？RFM是一个收集维度，个人PC、手机、平板电脑是另一个终端场景维度，PC能更好地收集R，手机能更好地收集M，这样就可以通过场景的不断变换来收集更多的数据。

现在，有一些终端的确可以收集以前收集不到的数据。以前，我们不知道一些数据的收集背景是不是移动的，但现在可以。用户做一件事情的时候是不是正在移动？他是不是在银泰百货里面？面对不同的场景，我们的框架也要相应改变。所以，现在做数据分析报告，最后的一个问题变成了："无线变了，这个报告的结果还是一样吗？你的报告应不应该也变一下？"

我认为，数据的本质就是还原，这是收集元数据的关键方法。如果没有这个概念，你就不知道未来你需要什么数据，就更不懂得什么是重要的数据，到最后只会产生越来越多的无从辨别的数据。一旦数据多到了连你的公司都没有办法处理的时候，那么其他会处理的人、公司和国家就会把你毫不留情地挤出市场。

"未来一定是国家和国家之间的数据大战，公司与公司之间的数据大战！"信息时代催生了海量数据的出现，这个世界上每时每刻都在产生大量的数据。此时，大数据已经不再是一个单纯的概念，而是像空气一样围绕在每个人的身边，每个人都是数据的制造者。也正是因为每个人都在通过不同的设备生产着数据，使得数据更多在"量"这个维度上不断膨胀，但是，"量"的单纯膨胀却对企业真正了解一个用户的需求产生了极大的挑战。所以，如何更好地识别各个设备的使用者是否为同一个人，如何更好地理解用户在各个不同场景下表现出来的不同需求，如何更好地理解数据融合后产生的价值，

将是未来商业中每一个企业都必须考虑的问题。

在不久的将来，随着O2O的深入和穿戴式设备的兴起，企业和企业之间必须进行更多的数据融合和交换，必须进行更多的跨行业的数据交流，这样才能更好地还原用户真正的需求，让用户在任何一个场景中都能够获得由数据带来的便利。

总而言之，更深化的数据连接使海量数据经过提炼更真实地还原了事实，也使我们运用数据科技去解码未来的需求成为可能。

[1] KPI（Key Performance Indicators），关键绩效指标，又称主要绩效指标、重要绩效指标等，是衡量管理工作成效最重要的一个指标。——编者注

[2] RFM模型（Recency Frequency Monetary），该模型通过一个客户的近期购买行为、购买的总体频率以及花了多少钱这三项指标来描述该客户的价值状况。

BIG DATA
The Business Revolution
数据化思考

CEO们关心哪三个数据

我在面试数据分析师的时候，必然会问他们一个问题：“假如我是一家知名电商的CEO，而今天是星期一早上9点钟，请你给我提供三个数据指标向我证明在过去的一周里，企业运营得一切正常，可以让我踏实下来。你认为，会是哪三个指标呢？”

绝大多数应聘者对这个问题的回答比较一致：第一个是流量；第二个是交易量；第三个是其他，这个其他包括转化率、交易额等。

当他们这样回答完后，我会反问他们：“刚刚我问的问题，你真的听清楚了吗？”

这时候，有人会回答说，我听清楚了，而答案就是这三个数据。

往往这个时候，我会提醒应聘者说：“请注意，我要的数据是给CEO看的，而且还是顶级电商的CEO，而且时间轴是周敏感数据。”面试进行到这一环节，我就发现大部分面试者根本听不懂“CEO”的含义。事实上，既然是CEO，就意味着他是公司里的最高领导层，那么给他看的东西明显要与其他人不同。

在这个例子中，我们会发现绝大多数应聘者很少会换位思考。也就是说，事实上，他们都是从自己的角度来思考，而不是以一个数据分析师、一个要给CEO汇报三个数据指标的分析师的身份来思考问题。

那么，什么是以数据分析师的身份来思考问题呢？通常来说，在我问出问题时，作为数据分析师的你首先要想的是CEO会关注什么数

据，是长期的，还是短期的？是风险最大的，还是风险一般的？或者是最近发生了什么事情？以及给CEO提供的数据要有什么注意事项，等等。

所以，我要再问问应聘者：“当你坐在面试桌对面给我答案的时候，有没有想过在星期一的早上，这家知名电商的CEO真正想看的是什么？”

再想想这个问题，你到底有没有真正听清楚“CEO”、“知名电商”、“周敏感数据”这些关键词？CEO要的是“踏实”——他听完了就可以安心地吃早饭了。

在面试时，如果面试者不对这几个问题进行询问就贸然回答的话，满分是10分，我只会给5分。因为这个问题里面本身就有很多问题，比如，什么是踏实？踏实是一个概念，你不问清楚“踏实”的含义，就给我三个指标，无论如何都是错的。

在正常情况下，首先不要急于回答我提出的问题，而是先问清楚什么是踏实，切勿自己先做假定。以下，我们可以假定一个相对理想的面试场景。

你反问：“什么是踏实的状况？”

我回答道：“最近这家电商和另一家电商在打价格战，而它最近又新推出了图书类目，那么CEO自然最关注的是这些图书的业务做得好不好。”

你再问：“什么是好？是否基于每天来买书的新增用户和原有用户购书的数量多少？而且，CEO是希望更多地用书来吸引新用户，还是想通过图书业务的推广让现有的用户进行交叉购买行为？”

在这些思考结束之前，你绝对不能给出指标。因为，在没有解决一个问题的内涵之前，任意给出的一个指标，必错无疑。所以，我才会问应聘者到底听清楚问题没有。

在我做面试官的经历中，很多看似有经验的数据分析师，往往在我提出的问题还没有解释清楚时就抢着作答。绝大多数人在思考不到一秒钟的时间里就给出了答案，而这一秒钟的答案，我可以确定他们根本

没有听清楚我的问题。

通常这个时候，我会再给他们一次机会，问他们："刚才你给我的这个答案，如果我给你满分10分，你会给自己打几分？"而此时，大部分人只会打6～7分。

当应聘者给自己打7分时，我会反问："另外3分丢的原因是什么？"他开始反思，说自己刚才给的可能并不是CEO想要的指标，因为他对这家电商的近况不是很了解……

接下来，当我再反问："刚才我的问题是'假如我是一家知名电商的CEO，今天是星期一早上9点钟，你给我三个数据指标向我证明在过去的一周里，企业运营得一切正常'，你听清楚了吗？如果你确认自己清楚了，能请你再给我一次答案吗？"

这时候，聪明的人不会再用一秒钟就给我答案了，而是重新思考，开始问问题，再给出答案。这时候的答案，当然会比第一个答案要好得多。

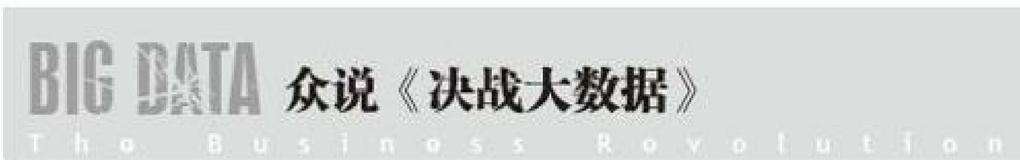
最后，当我再问他："现在，10分满分你给自己打多少分？"此时，他们自己给出的分值通常都会高一些。至此，我的面试也就结束了。

事实上，关于这个问题，我根本就不关注打分的结果。当然，如果评价是10分，那就不用面试了，因为在没有仔细考虑过答案的时候就自信满满地回答，这种人必然无法承担做数据分析师的责任。虽然，自信是对的，但是思考更重要。

作为一名数据分析师如果你不把自己的分析与当下结合，是没法进步的。

有趣的是，当我把这个问题贴在网上时，还是会有很多人追问我答案是什么、CEO关心的到底是哪三个数据。这时候，我真的很想说，答案不是结果，方法才是。

03 “活”的数据才是大数据



刘 星 红杉中国董事总经理

在中国，可能没有其他人比品觉玩的数据多，更没有人像他一样玩转了大数据。在这本书里，品觉以他特有的通俗易懂的语言风格，讲活了大数据，并结合阿里巴巴的实践，揭示了数据化运营和运营大数据的实战秘笈，灰常赞！

无论是企业还是个人，无论我们承认与否，在大数据时代，我们周围到处都充斥着碎片化的数据。为此，我们很容易迷失在海量的数据之中。

我们盲目地进行大数据投资，收集越来越多的数据。但是，令人沮丧的是，这些数据却是“死”数据。那么，什么是“死”数据呢？在我看来，“死”数据就是单纯存储在数据库中，无法进行分析和使用，并且不能够产生价值的数据。

那么，如何激活这些“死”数据，让整个大数据“活”起来，并成为企业运营的牵引力呢？答案就是：把数据用起来！

大数据的真正价值是将数据用于形成主动收集数据的良性循环中，以带动更多的数据进入这个自循环中，并应用于各个行业。举个最简单的例子来说，现在的很多网站都有推荐功能，很多推荐出来的东西，不论是音乐、视频，还是商品，都

可以让用户来选择“喜欢”或者“不喜欢”（当然，不点击未必代表不喜欢）。这样一来，企业就可以通过用户的选择基于计算机后台的算法为用户重新推荐，这就变成了一个循环——从基于已有的数据进行“分析—推荐—反馈—再推荐”的过程。当然，自循环还远不止这样一种形式。多样的自循环方式打开了大数据之门，而进入这个循环的关键就是，从解决问题出发。

目前，企业对于数据应用仍然持有半信半疑的态度。多年来，企业因无法建立数据收集的循环，致使其运营数据更多地建立在直觉的判断和分析基础之上。当面对周围海量的消费者数据时，充满了危机的大数据更难为企业的运作提供清晰的思路。无从下手，成为企业面临大数据时代的核心短板。这时，如果没有找出相关的关键解决方法，企业就会在由海量数据构成的新兴市场中错失发展的良机。

在数据的自循环中，有两个核心的关键点：一个是“活”做数据收集，另一个是“活”看数据指标。所谓“活”做数据收集，就是指企业不要局限于只收集自己用户产生的数据，还要把“别人”的数据收集过来进行综合分析；“活”看数据指标，就是指企业不要局限于已有的数据框架，而应该结合用户需求的不同场景来灵活应用收集到的“活”数据。

“活”做数据收集，抓住相关性

> “活”做数据收集，就是要跳出既定思维的框架，从相关联的行业和业务中去收集能够为现在所用的数据，找到能够更好地佐证企业现有业务决策和发展的数据。

关于数据收集，中国研究大数据的著名学者、电子科技大学的周涛教授曾分享了一个非常重要的观点。他说，做大数据收集不外乎两方面，一方面是“自己用”——用其他外面的数据来增加自己手上数据的精准度，为我所用；另一方面是“给别人用”——把我的数据贡献给外面很需要我的数据的人，从而

提高他的数据的精准度。

数据分析实例

> 亚马逊的数据收集有一个很经典的例子，恰恰体现了这一点。在很多年前，亚马逊就主动去收集用户的IP地址，然后从IP地址破译出用户所处位置的附近多少公里内是否有书店。工作人员从收集到的数据中了解到，一个人是否选择在网上买书，很重要的原因是他的附近有没有书店。

亚马逊的例子，是主动收集数据，即通过收集一个外部数据，来帮助自身判断线下是否存在潜在的竞争对手。

企业在做数据收集的时候，并不总是能够直接收集到所需要的关键数据，这时候就需要变通一下了。我曾经在一个外贸B2B平台工作过，当时很多人不知道我是如何找出新品种的趋势的，我也没有对任何人说过我的招数，有的人甚至以为在我背后有高手相助。而事实上，我之所以发现了新的品项，是通过追踪美国垂直行业电商的广告关键词变化来实现的。

中国的生产商跟美国消费者的距离如此远，我们如何才能收集到美国消费者的数据？一般来说，是不可能的。有一批厂家是从中国海关购买数据，但是海关的数据是滞后的，无法让企业及时获取行业的发展趋势。不过，这种趋势就在美国采购关键词的那些经理手上，而这个数据是可以从国外网站搜索到的，于是答案就这样到了我的手上。

在美国，这些搜索引擎营销（SEM）专家会为公司设计购买关键词的方案，而我就主要关注这些人的数据，看他们换了什么关键词，以了解美国的商品发展趋势。这样，我就可以很快地告诉类目经理，美国现在流行什么样的东西，而这就是利用外部数据来帮助自己做决策的方法。

以鞋类为例，我的做法具体来说，先是观察在美国做得最好的几家鞋类网站买入了什么关键词、变换了什么关键词、有什么关键词是常态的以及有什么关键词是在季节更迭的时候才买入的。在观察到这些关键词之后，我放到Google中去观察关键词的增长趋势，去eBay看看这一款鞋子有没有交易及价格的变化。在知道eBay有交易之后，我再放到国内电商平台中搜索这个产品。如果没有出现相应的搜索结果，就意味着这可能是一个良机。

所以说，做大数据收集，有时候需要更多的灵活变通。从上面可以看出，亚马逊的案例的确经典，因为它找到了消费者购买决策链条中的一个关键点；而B2B寻找关键词的重点在于如何观察第三者数据。每个人都知道在收集消费者数据时最好是观察直接用户，但如果没有这个数据，你观察什么数据？答案就是，去观察行业内对这个数据最敏感的那些人。

生活中其实也有这样的例子，我曾经给我的朋友分享了一个很简单且有趣的例子。如果你想知道香港的某家酒楼生意好不好，你问问门口卖报纸的人就知道了——香港人喜欢去喝茶的时候买一份报纸。

其实，这个规律不是我观察到的，而是香港税务局发现的。香港税务局如果担心酒楼对营业额虚报的话，就可以通过直接去查卖报纸的商家卖了多少份报纸来判断，这是一个非常有趣却很实际的灵活收集用户数据的案例。

“活”做数据收集，就是要跳出既定思维的框架，从相关联的行业和业务中去收集能够为现在所用的数据，找到能够更好地佐证企业现有业务决策和发展的数据。而“活”做数据收集的一大好处，就是能够规避现有数据框架的弊端，更好地反映用户的实际需求和企业的实际情况。媒体常用的“克强指数”是“活”做数据收集非常好的案例，通过耗电量、铁路运货

量和银行贷款发放量三大数据的结合来观察经济的发展，有助于剔除GDP统计数据中存在的水分。

“活”看数据指标，动态地使用数据

把数据激活，从静态数据变成动态数据，必须要用场景来验证。

我们不仅要灵活地收集数据，而且还要注意到，数据收集只是第一步，如果不让数据“活”起来，仅仅是把我们收集的数据简单堆砌在一起，是没有意义的。

用场景验证

在注册某个网站时，消费者常常需要填写自己的性别。我填写的答案是男性，但是如果你分析我的购买行为时会发现，很多时候我的账户在告诉网站，这些商品的目标用户并不是我自己，因为我也会为我的妻子和父母买东西。

当收集到的这些数据不能为企业所用时，企业就永远不知道关于我的这个数据原来是不准确的。这些数据好像是准确地描述了我的性别，但是却不能很准确地描述我的搜索和购物行为。

在梳理阿里巴巴的数据时，我们惊讶于为什么阿里巴巴会有18个性别标签。乍一听，很不可思议。你肯定会想，阿里巴巴是不是在闭门造车，是不是自己凭空造出了这么多的性别？

事实上，每一个性别表现都并非看上去那么简单，因为它的分类是基于用户在不同场景中不同表现而做出的。这就揭示了一个问题，我们每个人都不会只呈现出简单的一面，比如在安静时和在人前时，我们就会表现出不一样的自我。不同的性

别标签其实就是应用了这一原理——同样的人在搜索商品时可能会表现出不一样的行为特点，而这些不一样的行为就是我所说的场景，结合场景应用数据就是“活”用数据。其实，有多少个性别标签并不重要，重要的是如何让用户在不同的场景中获得更好的服务。

亚马逊一直在自己的商业活动中应用这个理论。多年以来，亚马逊就使用动态模型——相对于“历史的你”，它的推荐更相信“现在的你”。所以，它相信今天登录网站的你有什么需要与兴趣，比起历史的“你”来说更重要。

我们不妨扩张一下思路，假如你在一家网站买了一份保险，又买了一件衣服。这两件事情，哪一个能更多地证明你是“本我”，而哪一个能更多地证明你是“被别人影响的我”?

推荐系统的运作原理是通过推荐一件商品去影响你的购买决策。一般来说，你要为自己做一个很长远的购买决定时，会综合考虑一些因素，这时的购买决策会跟你本身的性格产生紧密联系。比如，你的性格是不爱晒自己的东西，不爱很刺激的活动，那么你就不会买一辆跑车，这是你深思熟虑之后做出的决定。

总的来看，对于用户购买行为的场景，我们大都无法判定它的产生到底是因为长期积累的性格多一点，还是受今天心情的影响多一点。而且，前提是假定，别人没有用你的账号购物。如果今天是你的女朋友借了你的账号去购物，这个系统可能就无法做出正确的推荐了。

把数据激活，从静态数据变成动态数据，必须要用场景来验证，静态数据是没有用的。理所当然，企业很容易一开始就假定“你告诉我身份证里你是男的，肯定就是最准确的”。从表面上来说，企业凭借用户自己填写的数据来做决策，会更

精准，但结果却相反。在做了很多测试后我们发现，动态的观察数据反而更能准确地反映出用户的信息和需求。

但是，如果用户在做一个比较重要的购买决定时，真实性别、真实职业以及一些描述个人的静态数据对于企业来说就更具有参考性了。对消费者来说，购物就是决策，凡是决策，都是要对比、细分、溯源、看趋势。所谓的冲动消费是“什么都不管了”，直接拍下，那就要分析购物当天的场景了，或许那天是“双十一”？

别再死盯着ROI

从公司价值的角度来看，我在B2B外贸公司得到了一个重要的经验——如果企业不把数据活用起来，只是死盯一个数据的话，企业的运营必定不会理想。

当时，我所在的公司买入广告关键词时，一直用的重要考核数据是投资回报率（ROI），即每个广告关键词能给公司贡献多少交易量。如果我们灵活使用数据，会怎么样？

我当年就对公司的管理者提出了一个建议：把一些使用得很好的广告关键词（排行榜前 30%）拿出来，调查根据这些关键词登录网站购物的消费者在三个月后是否还会来买东西的问题。

事实证明，这些ROI指标很好的广告关键词，有一半只吸引了用户的一次购买行为。因为，这些公司买的关键词都是“最便宜的××”，这些关键词对某些消费者来说有着很大的吸引力，但是并不总会产生巨大的效用。

也因为这样，企业吸引来的用户都是一般消费者，却不是真正做批发的人。企业管理者完全不知道当时掌控ROI的部门其实就是在做赔本生意。

所以，机械式地使用数据的结果就是，企业会因为“我们就看狭义ROI”死得很惨。ROI只能体现广告关键词在当天使用的效果，并不能反映出这个关键词以及其吸引来的用户对网站的长期价值。当时，公司把所谓的自己最“好”的广告关键词删掉之后，管理者发现3个月内企业的业绩没有受到任何影响。

“活”用数据，就是你是否能看出这个数据本身的局限是什么。一方面，是我们的数据为用户体验改善了什么；另一方面，企业在使用数据时，对活数据的运用解决了什么问题或者开创了什么机会。要牢牢记住，活用数据很重要。

“活”的数据是“活”用数据的精髓所在。《孙子兵法》曰：“兵无常势，水无常形，能因敌变化而取胜者，谓之神。”这句话在数据应用方面同样适用。如果企业能够基于场景和相关的“活”数据将数据应用发挥出最大的价值，那么新的商业模式的开创也就会在不远的将来成为可能。

BIG DATA
The Business Revolution
数据化思考

别再做“碰巧游戏”

在开始之前，我先讲一个有趣的故事。有一天，全球最大的职业社交网站忽然发现从雷曼兄弟公司登录的来访者多了起来，但是并没有深究原因。第二天，雷曼兄弟宣布倒闭。原来，前一天雷曼兄弟的员工都到职业社交网站找工作了。

Google宣布退出中国的前一个月，我在LinkedIn上发现了一些平时很少露面的Google产品经理在线，这也是相同的道理。

试想，如果LinkedIn能够针对某家上市公司选择性地分析数据，会不会挖出其中潜在的商业价值呢？我相信，现在很多网站还停留在“注册用户数”这样简单的数据收集上，绝对不知道还要收集上述有价值的数据信息来使企业获得发展。

好的分析师相当于市场解码器

我常常会说，今日的数据分析师，需要有点军师的味道——从枯燥的数据中看到解开市场发展密码的本事。简单来说，就是具有商业意识的数据分析师，如果监测到网站上拥有6岁之下孩子的用户群在增加，那么他基本可以预测出奶粉的销量也可能呈现上升趋势。

从更深一层来看，和传统卖场一样，网站中商品丰富性的作用大不相同。有的产品为了赚钱，有的产品为了促销，有的则是用来吸引流量，这也就解释了不同商品在网站上摆放的位置不同的现象。同理，只有具备了商业敏感度的数据分析师，才会懂得用什么数据来驱动公司实现经营目标，而绝不会单凭交易量就决定商品策略。

评估一个新生电商平台的竞争力，你会观察什么数据

交易量？不对，因为平台是双边市场，不同于单边的B2C卖场，除了买家流量之外，对于一个需要活力的商业平台来说，每天有多少新的优质卖家加入进来也至关重要。比如，他们卖了多少东西和卖了什么东西，这对新生的电商的竞争力来说都起到了不可或缺的作用。因为，此阶段商业竞争最核心的影响因素是人气，而非实质交易量。如果新加入的卖家卖不出东西，只是老卖家的交易量在增长，即使最后的交易量每天都呈增长趋势，也无法极大地提升其竞争力。

换个场景看，一家刚踏入市场的B2C和已经占领大部分市场的B2C，它们的关注点会一样吗？当然不可能，因为前者看的是流量，赚的是人气；但流量对于后者的意义就没有那么大了，因为它们更关注交易率、转化率和回头率。

组装你的数据仪表盘

我把数据信息的梳理过程，理解为数据世界里的“仪表盘”。比如，在开车时，如果水温过高，仪表盘就会亮灯提示；同样，网站的关键信息也可以监控，在电子商务交易中，行为数据和商业模式之间的逻辑关系组成了“仪表盘”——好的仪表盘，对于商业活动中的好现象和坏现象都会有敏锐的反应。

为了方便理解，我想出了两个名词：前端行为数据和后端商业数据。前者指的是访问量、浏览量、点击率及站内搜索等，是反映用户行为的数
据；而后者更侧重于商业数据，比如交易量、投资回报率、客户终生价值［LTV(Life Time Value)］。

分析数据　　“走火入魔”的人会明白，数据就像散布在夜幕里的繁星，它们彼此之间布满了关系网，只要轻轻按动其中一个亮点，就会驱动另外一个亮点产生变化，而这个动态的网，就是“仪表盘”。

将行为数据与商业数据对接起来时，我们大家可能都会比较关心网站的用户群，以下就是一个相关的例子。

某天，某网站发现自己的前端注册量增加了不少，访问量也提高了，但交易量却未见大幅度的提升。原因是什么？估计数据分析师们每天都在思考这个问题，那么，在这里我们就将这个问题进行分解，以寻找相应的答案。

处在网站前端的人只知道点击量等数据，比如谁一直在重复购买？谁影响了5%～15%核心用户群的购买行为？谁在给网站做正面或负面传播？但却很少有人会关注后端的商业数据。

如果操作网站后端交易环节的工作人员，只知道卖东西，却不知道提出这几个问题：一个客户登录网站平均停留时间是15分钟，还是30分钟，这对将来重复购买的影响大吗？一个客户登录了网站社区和没登录社区，对产生交易量有影响吗？此时，却很少有人会关注前端数据。

你发现这中间的问题了吗？行为数据与商业数据没有对接起来！

于是，前后端数据形成割裂，没有人知道它们的内在关系；作为网站的决策者，不知道网站核心用户群的行为特征，也不知道怎么去刺激核心用户的增加，更不知道从一个用户登录网站之后到离开哪些环节需要疏通的话，不死就真的是万幸了。

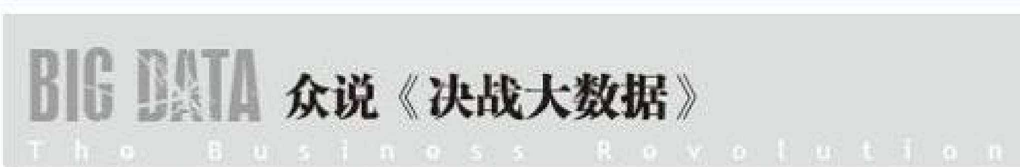
当然，这个例子也只是管中窥豹。一个平台运营商，反映用户行为的前端数据与后端的商业数据千千万万，卖家和买家也是万万千千，但是在了解前端哪一个数据对整个网站后端的交易量有最大影响后，就只须针对这个前端数据猛出力，必然会刺激后端数据的增加；反过来，后端哪一个交易数据比较高，弄清楚是从哪一个渠道收集的、主要贡献用户是谁，网站的产品设计就要向他们倾斜，如此才会使渠道前端“转化率”等关键数据得到提升。

如果一个网站的核心用户群每月以10%的速度在增长，不火也是怪事。

遗憾的是，今天许多电子商务公司，每天都在做“碰巧”游戏：今天推荐A家的产品，明天撤下A家的产品；今天做低价促销，明天又做线下活动。这些决策的改变，没有“仪表盘”的指示或良好的监控，都是闭着眼睛在“碰巧”。

世上哪有这么多“碰巧”，还是快睁开眼睛，认真看路吧！

04 无线数据，大数据的颠覆者



张 涛 大众点评网**CEO**

作为一个企业经营者，说起大数据，最怕只见数据不见价值；作为一个互联网产品老兵，最焦虑于数据挖掘和分析脱离用户场景。本书好就好在实践了作者书中提出的观点，立足应用场景，聚焦如何让数据产生实际价值。因其实战性和创新的角度，值得所有关心大数据实战者认真一读。

2013年被称为无线元年，智能手机的出货量超过了3亿部、4G牌照的发放、微信等应用的快速崛起、阿里巴巴移动平台“ALL IN”的发布，这些现象和事件无不意味着移动时代的到来。移动数据在不断改变着人们的生活，也在不断改变着数据信息的全貌。现今的大数据环境，与之前相比，变得更为复杂，也给数据的分析带来了诸多挑战。

让我们再来回想一下第2章描述过的一个很多人都有可能碰到的场景：

> 早上在路上你发现了一个路人身上穿的一件T恤很好看，于是，在进了办公室的第一件事情就是打开电脑，登录网站搜索这件衣服。在还没有挑选好的时候突然被老板叫去开会，开会的时候觉得有些无聊，就掏出手机继续搜索，后来被手机应用上的一个手机活动吸引了，最终买下了一部手机。

这个看似简单的场景，从数据分析的角度来说却非常不简单。如果按照过去的做法，我们是用PC的视角去分析这一数据的话，那么就会发现这个用户搜索了T恤，浏览了T恤的几个页面，但是在几个小时后却买了一部手机。这不仅仅会让数据分析师变得困惑，更让整个基于浏览和购买的推荐系统产生疑问——此种行为会将T恤和手机这两个内容产生一种相关，但事实上这两类东西显然是不相关的。

这时候，我们需要的是一种多屏思维——需要考虑到现在用户使用互联网的多场景问题。现在，多屏可能会包括多台电脑、手机和pad，而在可以预见的不远的将来，可能会有更多可穿戴式设备的出现，比如用智能眼镜来搜索一件衣服，用智能手表录下的一段语音来搜索某一种好吃的零食。当多屏变成一种常态时，不管是数据分析师的分析方法，还是推荐系统的推荐算法都必须多屏化。因为当多屏时代到来后，一切数据都在发生变化……

无线数据正在将整个数据变成“噪音”

无线时代到来之后，一切都发生了变化。原本可以收集用户完整行为数据的网站变得不再可靠，而数据还原用户行为的方式也发生了改变。

在探讨无线数据让数据产生“噪音”的问题前，我觉得还是有必要稍微谈一下关于数据收集的方法。PC时代的互联网，数据收集是基于浏览器的。从20世纪90年代开始蓬勃发展，在这20年左右的时间里，整个互联网已经有了比较成熟的数据收集方案——基于cookie来进行用户身份的识别。如果是一个拥有用户体系的网站，那么就可以使用用户ID与cookie的关联性去识别用户的行为。在PC时代，用户在使用浏览器时，cookie通常都是长期存在的，而且对于大部分用户来说很少会在意它的存在，因为Cookie对用户的意义不大，对网站的意义才大，这是由于网站可以通过它长期地收集用户的行为数据。一个电子商务网站能够知道用户看了什么、什么时候看的、看了多长

时间、结果买了什么……这一系列的场景它都能够了若指掌。这种收集数据的方法，可以让网站能够更好地去识别用户，而这也是研究用户行为的第一步。

可是，无线时代到来之后，一切都发生了变化。原本可以收集用户完整行为数据的网站变得不再可靠，而数据还原用户行为的方式也发生了改变。那么，为什么无线的出现会让这样的情况发生呢？

在这里，我们简单讲解一下无线数据收集的原理。无线基本上分为两种形式，一种是**WAP**，另一种是**APP**。对于WAP来说，整体的工作原理和PC差不多。在移动互联网早期，用户可以通过手机的浏览器来访问专门为手机设计的页面，当年最有名的WAP网站莫过于移动梦网了。

而现在移动互联网的主流应该是APP，是基于手机端的应用来运作的。APP的发展速度非常快，尤其是在最近几年。随着iOS和安卓的高速发展，整体的APP用户也在快速地增长，但是，数据收集的技术却并没有像PC那样成熟。APP数据收集的方式包括两种：一是收集用户联网时请求服务器的记录；二是在将用户的行为数据记录下来之后，适时地传给网站。但不论是什么样的收集方式，无线数据最终的表现是在没有账户体系的情况下，和**PC**的用户行为完全没有办法进行关联，这也就意味着用户的数据出现了断层。

如果只收集无线终端的用户数据是不是就不会出现“断层”的问题呢？答案也是否定的。在PC时代，虽然浏览器各有不同，但是基于cookie的整套数据收集体系很成熟，不同的浏览器不会对数据的收集造成很大的影响，但是无线APP时代就完全不一样了。APP对用户的识别是基于手机的机器码，但是这一识别标准在不同的操作系统上也存在差异。这种差异主要体现为以下两点：一是操作系统本身的差异；二是同一操作系

统上的版本的分化。操作系统的差异比较好理解，现在主流的操作系统iOS、安卓、Windows Phone就很不一样，他们的安全策略也非常不一样，比如iOS的封闭式系统就会严格控制外部收集者的介入，限制其获取自身用户的信息。而版本的分化会对数据收集造成更大的挑战，比如iOS6的操作系统可以允许应用程序获取手机的MAC地址，而到了iOS7就无法获取了，应用程序就只能获取广告识别码，但用户又可以限制应用进行获取，这就导致基于MAC地址识别用户的方法在用户升级操作系统之后就完全失效了。安卓在这一方面的表现更加夸张。安卓是一个开放的系统，不同的厂商，甚至其他组织都可以修改它，这就使得在一部分手机上可以使用的方法在其他手机上就行不通。典型的例子就是通过手机扫描二维码打开应用，用小米手机扫描二维码就能打开应用，而三星手机也许就不行。

对于APP来说，还有很多APP会内嵌网页，这个时候往往会收集到应用本身的数据和网站浏览的数据，这就使得数据变成了两份。如果没有将一个应用内的识别用户的标记很好地进行传递，那么即使在手机端，数据也会变得不完整。

再退一步讲，即便上面说到的两点都没有对数据收集产生影响，APP的版本更新也会对数据收集造成障碍，比如一个APP开发的时间长了之后，总会更新多个版本，而现实中也总有不愿意更新版本的用户，这就使得数据收集的结果会存在多份的情况，而不像**PC**那样只产生一份数据。比如，对于手机淘宝来说，最新的版本可能是4.0，但是手机预装的版本可能还是3.4，在这种常见的情况下，就要求数据分析师可以从更广的视角去看待数据收集的问题了。

PC数据与无线数据的关系

保证**PC**和无线两份数据的完整，通过用户体系将两份数据关联起来，就可以在分析的时候用彼此的融合来还原用户的行为。

在两份数据同时存在的情况下，就需要去衡量PC和无线的关系。我的看法是，保证PC和无线两份数据的完整，通过用户体系将两份数据关联起来，就可以在分析的时候用彼此的融合来还原用户的行为。

首先，从数据本身来看，PC和无线两份数据由于存在技术上的差异，本身就是分离的，这就意味着数据仓库需要用额外的空间将无线数据保存起来，而不能单纯地将无线数据混入PC的数据里面。而且，无线数据中还存在APP性能的数据，比如后台启动、程序崩溃等，所以分开存储是十分有必要的。对于不需要还原用户行为的数据，比如页面整体的点击率、用户访问时长等基于应用本身的数据，则可以通过无线数据来分析，这样效率更高。

其次，对于数据收集和管理的策略而言，必须要有一套账号体系能够使原本没有关系的数据发生关系。由于技术不同，PC和无线数据天生就不发生关系，要让关系发生的唯一办法就是人，或者说是基于用户的账号。这对于一些有账号系统的平台，比如腾讯和阿里巴巴，这不会是多大的问题，但是对于不依赖账号系统的平台来说，想让数据之间发生关系十分艰难。为了让用户的行为能够串联起来，现在很多应用都要求用户进行登录，这样即便是用户换了手机之后，用户的行为仍然能够得到延续，而企业也会清楚此用户是老用户。否则，在换手机如换衣服的今天，可能所有用户手机的生命周期都不到两年，那么企业就永远都不会知道谁是自己的老用户了。

多屏思维下的电子商务

探索到的无线数据场景为清晰的数据分析打下了基础，从而能够更好地还原用户的行为，以便分析在多屏环境下的电子商务。

面对无线数据，要解决的主要问题有两个：一是做到高效准确的收集；二是培养数据分析师的多屏思维。

所谓做到高效准确的收集，就是需要建立一套面向多操作系统、多版本、自动、可灵活配置的用户行为获取系统。通过这个系统，开发工程师不再需要采用人工写代码的方式来收集数据，从而也就保证了数据的纯净无噪音。但是，这套系统还需要应对一个问题，那就是可能出现的因操作系统的变化而使用户识别失效的情形。因此，企业需要能够建立自己的用户识别方式，能够建立独立于设备号之外的识别代码。同时，这套系统还需要做到多系统的适配，要保证同一个版本在安卓和iOS上面的数据具有一致性。

数据分析师多屏思维的培养，则是一个长期的过程。我的方法是先派出精干的数据分析师小团队去探索无线数据的场景，因为即便像阿里这样的大公司，无线数据的分析也是一个全新的领域，所以需要有分析师去探索。经过探索之后，我们就能够发现无线的各种“坑”（问题），为清晰的数据分析打下基础，从而能够更好地还原用户的行为，以便分析在多屏环境下的电子商务。

最后，在数据存储领域，也需要有更合适的模型框架来应对海量的无线数据。在无线出现之前，数据仓库都是基于PC的思路，按照主体来建立的，比如用户属性、浏览行为、订单中心等；在无线出现之后，很多数据仓库工程师则将无线也作为一个并列于用户属性或者浏览行为的主体进行保存。

其实，这种做法的误区是没有真正理解无线的含义，无线应当是作为一种横向的基础模型穿插于各个垂直的主题模型之中的，如果将无线作为一个独立的模型，在进行用户行为还原和用户分析的时候则会变得很困难。

古往今来，每当颠覆者出现的时候，机会也会随之出现。无线互联网是当今商业发展不能避开的鸿沟，此时，企业应该做的是在颠覆者行动之前，提早地做好准备和规划。在无线时

代和多屏时代，企业需要设计更好的技术方案来让自身更好地识别用户，用更好的产品场景让用户享受到多屏的便利，及早地培养数据使用人员的多屏思维方式，及早地规划面向无线场景的数据底层设计方案。

从数据化运营角度来看，云是计算，端是应用。“云＋端”的模式使数据科技的价值得到了极大的发挥， 就像“INTEL inside”一样。从运营数据的角度来看，端是传感器，云是“储存＋管理”，而“多屏＋移动端口”就是未来数据创新的孵化器。

在未来的商业大战中，积极备战无线领域的公司，将会在移动互联网的场景下独占超越其他公司的优势，并享受更为丰厚的收益。

BIG DATA
The Business Revolution
数据化思考

样本的偏见

假设，你想买一款新手机，于是去询问三位朋友的建议——到底是买iPhone，还是买三星。如果我们把这种经历当作一次抽样调查的话，我确定，它绝对不可能告诉你真实的结果。

我们在做抽样调查的时候，样本往往从一开始就有了最常见的“偏见”——谁愿意反馈我的问卷，谁的建议就被反映在调查结果中；而不愿意给我反馈的人，他的意见就不会反映在调查结果里。有趣的是，被我们忽视了的那些没有反馈的样本，很多时候却是我们最需要知道的结果。

比如，今天淘宝做了一项调查，问：“你爱淘宝吗？”在收到1万份回复为“爱”时，我们的调研报告就得出了“大家都爱淘宝”这一结论。但事实上，我们总共发出了10万个样本，这一结果说明有9万个样本对这个调查保持了沉默。此外，我们在做调研时，会向被调查者赠送礼物，有些人是因为收到了礼物才觉得应该“说点好话”。在了解到这样的情况后，你还会轻易得出“大家都爱淘宝”的结论吗？所以，当我们分析一份数据的时候，一定要问自己：“样本的背后有什么是我们不知道的。”

我们经常会看到易观、Comscore、艾瑞等机构公布某一事物的调研数据。在看到这样的数据时，我常会问，他们的样本是怎么挑选的？据我所知，他们有一个第三方监控软件，它会以数百元每月的费用“购买”样本，并把监控软件置于后台。试想，如果你想做一个高端样本的调研，那么这种做法可行的可能性很小，就算有可能，结果也不会准确。因为他们即使出1 500元每月的资金也无法获取准确的数据。缘由显而易见，让月薪为5万元以上的用户把软件放在后台，然后监测自己

每天做了什么，他们自然不会答应。所以说，在这种取样结果中，注定有一部分数据缺失而最终导致结果不正确。

在现实生活中，我们往往不会去想采样的问题，这就造成我们对很多调查结果深信不疑，我们的决策就会有对号入座的嫌疑，而这种嫌疑非常常见。比如，某人告诉你说他身边的很多朋友也是天秤座，这是因为他心里能想到的朋友很多都是天秤座，事实上其他他未能想到的人也可能是别的星座。再举两个现实中的例子，解释一下样本的偏见。比如，某人说他开了2 000次的汽车从没有出现过意外。今天，他喝酒了，但那2 000次里从来没有喝酒后驾驶的样本，所以2 000次这个数据在今天是没有参考意义的，因为并不能证明他喝完酒之后还能安全驾驶。而如果他说有过200次喝了超过3瓶白酒，而且喝完之后开了50公里回家的经历。那么，这200个样本就有参考性了。

样本数量是否足够和是否平衡的问题，是另一个常见的“偏见”。某杂志对用户们偏爱的手机操作系统做了一项调研，结论是现实中的人更多地在使用安卓系统。但是，这本杂志的读者群主要是IT界人士——这个取样一开始就存在偏见，而抽样又是来自读者群，再加上IT界人士本身也比较倾向于接受安卓系统，所以样本的不平衡就造成了结论的失真。

另外一个影响样本的因素是时间。我们公司有一个360度调查，用于反馈员工对部门的评价。如果我想拿高分，我在调研开始前请大家吃饭，给大家培训……我做好所有调查中出现的项目。你要相信，短时记忆力是十分强大的，这就是一个典型的时间样本。这样的抽样，是典型的“似是而非”。那么，假如想消除时间因素对结果的干扰怎么办？其实很简单，你只要随便抽取一天来询问员工：“你的部门好吗？”这往往才会得出有价值的答案。

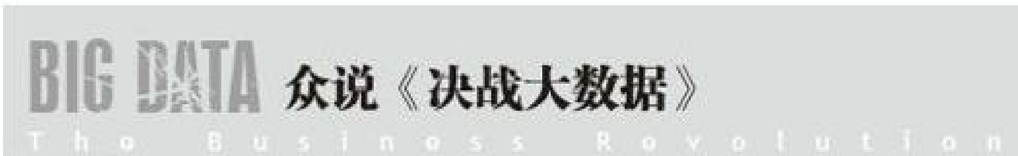
此外，偏见还会因为我们抽样时的询问方式出现。这里有一个很好的例子，Amazon的调研就很让人信服，它的问卷第一句话就是：“你的亲戚好友里，有没有亚马逊员工？”如果你回答“是”，那么很可能你的意见就不会在这份样本结果里出现。亚马逊的员工、亲友和普通亚马逊消费者的观念是很可能存在差异的。亚马逊的问法，使得调研出的数据更加精确，让有“偏见”的样本也存在了意义，因为它能区分双方的用户层次有多少不一样。

在调研过程中，如果这个样本在你调研前就是有目的的，那么这份有偏见的数据也是有价值的，比如可以通过交叉信息来验证结果是否准确。这在我们身边很常见，比如，你问三个朋友，我是买iPhone还是三星。但首先你得问：“你用的是什么手机？”如果他买了iPhone，却推荐三星，你就要追问为什么了。

同样，我现在去询问淘宝小二：“你会选择在哪里购物？”他们熟悉淘宝，肯定会受到影响。而如果对小二们调研，结果显示更多人愿意去京东、当当购买商品时，这就证明京东和当当确实不错——起码就某些类别来说是不错的。此时，有偏见的样本对企业仍然存在价值。

值得注意的是，样本跟大数据不同。大数据相信全量数据，而非样本；是分析得出，而不是抽样获得。

05 数据分类与数据价值，什么才是你的核心数据



程 杰 **Acxiom**（安客诚） 全球副总裁，数据科学家

大数据像铺天盖地的洪水一样涌来！由于近年来数字媒体、网络和移动技术的迅猛发展，数据的积累速度已对数据的存储、管理、分析和决策应用提出了前所未有的挑战。很多企业、政府、学校和研究机构为了能在大数据时代继续生存和发展都在重新定位和寻找新的方向。让我们高兴的是，作为一个先行者，品觉在阿里巴巴已经走出了一条大数据运营的路，并又在运营大数据方面总结了很多成功经验。他的《决战大数据》一书尤其为读者提供了“数据化思考”的模式和框架。通过用实例讲故事、作比喻，品觉打破了大数据的神秘，然而又能使读者产生丰富的联想，开动脑筋，真正理解大数据成功运用的要诀。感谢品觉为“大数据金矿”的探索开发作出的贡献。

在大数据时代，首先要做的是收集大量数据，但收集数据并非仅是把收集过来的数据放到硬盘里面那么简单，更需要的是对数据进行分类、存放及管理。不然就如同一个杂乱的储藏室——放东西进去的时候很轻松，等到要找东西出来的时候就会浪费大量的时间，甚至可能再也找不到。如今，数据在企业中已经充当了一种生产原材料的角色，而既然是原材料就要被估值。

从数据角度来说，估值就是通过不同的维度去思考数据的价值。只有基于对数据的分类和对数据价值的不同认识，才能去对数据做筛选。如果连有多少种数据都不知道的话，我们

很难描述数据被收集后到底如何进行筛选，也不知道哪一个数据更有价值，更不知道哪一个数据需要更安全的处理，甚至不知道在它的生命周期中做备份的策略是什么。

对于数据的认知，完全取决于我们是否拥有认知自己所拥有数据的能力，是否能够筛选出到底什么是我的核心数据，到底什么数据会被我们频繁地使用。

数据分类为什么如此重要

> 数据作为一种资产，不同的数据含金量必然不同，自然就会产生不同的价值。而且，就同一组数据而言，在不同的环境下甚至会呈现出不同的价值。

在大数据时代，每做一件事情我们都要知道确切的目的是什么。从自我的角度考虑，当我们给数据分类之后，真的就能更容易理解大数据吗?

“数据”本来就是一个既清晰又模糊的概念。之所以说“清晰”，是因为每个人对数据都有概念，而且每个人每天都会接触到各种各样的数据；而说“模糊”，则是指数据本身包含了各种不同的类型，能够产生完全不一样的价值。

至于数据到底应该怎么进行分类，则需要视不同的场景而定。就好比“人类”一样，如果拿人类和动植物来做比较，人类整体必然有不同于其他物种的特性，这时，“人类”就是一个非常“清晰”的概念。但是，如果要对“人类”进行更深层次的分析，显然需要对不同类型的“人”进行比较和分析。在很多场合下，可能还要对不同种族和不同年龄层进行分类和分析。

数据作为一种资产，不同的数据含金量必然不同，自然就会产生不同的价值。而且，就同一组数据而言，在不同的环境下甚至会呈现出不同的价值。就好比衡量黄金和铁这两种金属的经济价值，同样重量的黄金其价值必然远胜于铁；但是如果

用来做刀剑，那柔软的黄金显然就不是一种合适的材料；而黄金和铁的合金可能又能够在另一个场合有一定的应用。

对于数据而言，不同的场景也会产生不同的价值，有些数据可能会变成一个运营指标，能够让管理者更好地对业务进行决策；有些数据可以对未来进行很好的预测，而不同的数据进行有机结合之后又能够对用户做出更好的商品推荐。一个好的工匠必然对各种材料的性质及使用方法了如指掌，而一个好的数据从业者则必须要对各个数据的价值和稳定性洞若观火。

遗憾的是，并不是每个数据从业者都能够很好地对数据进行分类和估值。很多时候，我们可能只看到了数据的一些应用场景，却不能够更加深入地进行数据的管理和应用。

当我们将焦点放在一组数据“能产生什么价值”上时，往往会忽略其生产过程，因为我们的很多数据是经历了加工才产生这个价值的，而往往其中整个加工链条的中间数据并未被区分出来。就好像我们在吃一盘西红柿炒鸡蛋，好吃是结果，但如果你问西红柿给“好吃”贡献了什么，鸡蛋又贡献了什么，你可能就回答不出来了。

古语有云：“治大国若烹小鲜。”其实做数据也要有这样的感觉，尤其是在大数据时代，更是要做到精细化。还是拿西红柿炒鸡蛋这个例子来解释。

> 我们这次做的可能是这样的一盘，那么下次是不是还能保证做的口味完全一样呢？如果要做到口味完全一样，那我们应该怎么做呢？当然，你可以说这个过程已然成了一种习惯，是一种感觉，但是一旦需要做到标准化或者所有的人都要做到一样，与这样的“习惯”相比，就不如找出当中的规律了。
>
> 这个过程是一个定量分类和管理的过程，也是一个标准化的过程——多少西红柿和多少鸡蛋。如果再精细一点就要看哪里种

植的西红柿、哪一个养殖场的鸡蛋、西红柿的成熟程度怎么样、鸡蛋一个要多少克，等等，所有这些内容都要有精细化的规定。

将之应用到数据上，特别是在运用数据上，这样一个精密的过程就更加必不可少了。比如说，这次可能运用了一个模型，效果不错，这次用的这组数据效果也不错，那下次是不是还能保证有类似的效果呢？或者说，场景变换了还能保证类似的效果吗？这些都是需要去认真思考的问题。

数据分类的4大维度

在观察与分析数据中，我们要从中间抽象出来，更好地将数据进行归类和整理，从而更加清晰地识别出数据的价值。

今天，我们多数人把自己关注的焦点放在结果上，忽略了过程，而事实上这些过程其实也蕴含了数据管理的过程和数据分类的过程。有些关键数据必须做好保护，若这几个数据变了，或者被污染了，前面的价值也就无法保证了。有人肯定会问，管理数据是一个过程，能不能分解？能不能区分哪些元素是比较重要的？哪些是有代替品的？哪些是无可取代的？哪些是不可或缺的核心？其实，这些问题就包含数据战略。

所以，我们不可以含糊地应对，要从中间抽象出来，更好地将数据进行归类和整理，从而更加清晰地识别出数据的价值。

从数据分类的角度来看，可以将之分为以下4种。

1.按照是否可以再生的标准来看，可以分为不可再生数据和可再生数据。不可再生数据通常就是最原始的数据，比如用户在访问网站时，浏览记录会追踪用户的行为，如果当时没有被记录下来，就没有其他数据来还原用户的行为了。这个有点像拿着相机拍闪电，抓拍很重要，一旦错过，闪电就不可

能再重复刚才那一瞬间的光影了。因此，对于用户日志类等不可再生数据而言，必须要有很完善的保护措施和严格的权限设置。现在，很多系统都有备份多份数据的功能，理想情况应该是，因为磁盘损坏而造成数据丢失的案例应该越来越少。但是，因为系统升级失败和误操作等失误造成的数据丢失在各家公司都屡见不鲜，见怪不怪了。

可再生数据就是通过其他数据可以生成的数据，原则上，指标类数据的衍生数据都是可再生的——只要原始的不可再生数据还在，就可以通过重新运算来获得。不过千万不能因为“可再生”这个词语的存在，就对可再生数据掉以轻心。有些可再生数据是通过很长时间的积累不断加工而成的，是长时间从海量数据中计算出来的，比如对某个用户在数个月内的连续购买行为产生的规律，如果未做保护，虽然仍然可再生，但是再生的时间却会给企业带来问题。因为即便对于有顶尖计算能力的公司来讲，都可能是数日，甚至是数周、数月，而这个时间过程可能就会对公司的某一项核心业务造成毁灭性的打击。

对不可再生的数据而言，已有的数据要严格保护，想要但是还没有的数据就要及早收集。举个例子，很多电子商务网站是不关注客户在商品详情页面有没有做滚屏操作的。如果这一类型的数据没有被记录下来，企业就无从知道详情页的有效性。当商品页面进行改版，需要对此类数据进行参考时，就没有办法来获得相应的数据支持，最后能做的就只能是等待在页面上进行布点开发，等待数据收集到之后再进行决策，这就造成了决策的延误。

对于可再生数据而言，要及早做好业务的预判和数据处理的规划，这样一来，数据在需要的时候就能够快速地获得应用，我们把这一数据叫作数据中间层。

2.按照数据所处的存储层次来看，可以分为基础层、中

间层和应用层。从数据的存储角度来说，数据有很多层次。基础层通常与原始数据基本一致，也就是仅仅存储最基本的数据，不做汇总，以尽量避免失真，从而用作其他数据研究的基础；中间层是基于基础层加工的数据，通常也被认为是数据仓库层，这些数据会根据不同的业务需求，按照不同的主体来进行存放；应用层则是针对具体数据问题的应用，比如作为解决具体问题的数据分析和数据挖掘的应用层的数据。

在存储层这个层面上，最大的问题就是数据的冗余和管理的混乱。尤其是对于一些拥有海量数据的大公司而言，数据的冗余问题尤为严重，由此造成了大量的浪费。在大公司中，进行数据分析、开发、挖掘的人可能有数十甚至是数百人，这些人可能归属于不同的业务团队，为了满足不同的业务各自分析数据应用。这样一来，不同的人可能都从头开始建立起了一套包含基础层、中间层和应用层的数据，而彼此之间又没有合适的交流方式，也就造成了工作的浪费。那是不是应该把所有的数据进行更好的归纳或者管理呢？我认为这也不是绝对的。任何管理方法，无论是集中式管理，还是分散式管理，都各有利弊，而且人和业务多了之后，企业也很难进行集中式管理。我给出的建议是，基础层必须统一，因为这是最基本的数据，而且基本数据是原始数据。除了备份的需求外没有必要在各个场合保留多份数据。只要保证这个数据有良好的元数据管理方式，就能极大地降低成本。而对于中间层和应用层而言，则要视具体情况而定：如果公司的业务相对单一且成本压力比较大，则建议集中式管理；如果公司的业务量非常大，则可以由多个数据团队来进行分散式管理和应用，以保证基础层单位有最高的灵活性。

3.按照数据业务归属来看，可以分为各个数据主体。按照业务归属分类的意思就是，将数据按照不同的业务主体分门别类地进行归纳。就好像仓库一样，将不同的物料进行分类存放，可以提高其使用和管理的效率。按照业务归属分类的数据

在不同公司可能体现出不同的内容，在平台型电商可以分为交易类数据、会员类数据、日志类数据等。交易类数据是指平台型电商的订单流水，其中包含了买家、卖家在什么时间成交了什么商品；会员类数据记录了买家、卖家的身份信息，比如注册时间、身份证号码、信用等级等信息；日志类数据则更多的是指用户的行为，即哪个用户在什么时间段访问了平台的什么页面、点击了什么按钮等。

对于数据的分类主体，则要根据业务特点进行归类，并没有一个特别的硬性规定。总体的原则就是让数据的存储空间更少，分析及挖掘的过程更简单、快捷。

4.按照是否为隐私来区分，可以分为隐私数据和非隐私数据。顾名思义，隐私数据就是需要有严格的保密措施来保护的数据，否则会对用户的隐私造成威胁。用户的交易记录属于隐私类数据，对于一家有着良好数据管理机制的公司而言，通常的管理方法是对数据的隐私级别进行分层，数据从安全的角度可以进行两种类型、四个层次的数据分层。两种类型就是企业级别和用户级别。企业级别的数据，包括交易额、利润、某大型活动的成交额等；个人级别的数据就像是刚才提到的身份证号码、密码、用户名、手机号等。四个层次是对数据进行分类，分别有公开数据、内部数据、 保密数据、 机密数据。

当然，也有隐私数据保护得不好的企业，之前很多隐私泄露的案例都对用户造成了很大的损害。比如，某些网站几十万的开房信息泄露、数百万的密码泄露等都是类似的事故。随着拥有大量数据的网站和公司越来越多，数据安全就越来越成为一个核心的点——需要投入专门的人和专门的团队来进行数据安全的管理。而数据安全工作的推动，初期往往会受到一线员工的反对，因为任何一个安全系统都意味着已有的权限被收回，也会因为改变工作方法而降低效率。所以，拥有大数据的企业高管必须要关注数据安全，否则数据越大，对“恶

人”的吸引力就越大，最终用户和公司的损失也就越大。

数据的5大价值

> 能够辨别关系、身份的数据是最重要的。这些数据应该是有多少存多少，永远不要放弃。在大数据时代，越能够还原用户真实身份和真实行为的数据，就越能够让企业在大数据竞争中保持战略优势。

如果说数据分类的目的是为了对其有更好的认知和管理的话，那么，对不同数据价值的认知就是让数据更好地被应用的前提。对于数据来说，如果仅是将数据收集起来不进行运用，那除了占用存储空间、加大存储成本之外没有任何价值。但在实际运用中，也需要认清数据到底能够产生什么价值：有时候，同一组数据可能会在不同场合产生完全不一样的价值；有时候，单一的数据没有什么特别的价值，需要组合起来才能产生价值。

数据价值1：识别与串联价值

顾名思义，识别的价值，肯定是唯一能够锁定你目标的数据。最有价值的比如身份证、信用卡，还有E–mail、手机号码等，这些都是识别和串联价值很高的数据。在前面我多次提到的搜索T恤的例子中，电商网站识别“你”的方法就是你的登录账号。千万不要小看这个账号，如果没有这个账号，网站就只能知道有一些商品被用户浏览了，但是却无法知道是被哪个用户浏览了，更不可能还原出用户的购买行为。

当然，识别用户的方法不止登录账号一种，对用户进行识别的传统方法还包括cookie。所谓的cookie就是在你浏览器里面的一串字符，对于一个互联网公司来说，这就是识别用户身份的一个标记（当然也可以选择禁止被追踪），所以你就会发现你在搜索引擎上搜索过一个词语，在很多网站都看到相关的资讯或者商品的推荐，就是通过cookie来实现的。很多互联网

公司都非常依赖cookie，所以会采用各种cookie来记录不同的用户类别，单一的cookie没有价值，将用户登录不同页面的行为串联起来才产生了核心价值——串联价值。

然而，现代社会已经不再是一个单纯的“PC＋浏览器”的时代了。2013年，被很多人认为是移动互联网元年——用户接触互联网的方式正在快速地向“PC+手机+pad”的形式转移，互联网已经进入了多屏时代。在这个信息多元化的时代，单纯运用**cookie**来识别用户行为的方式已经过时了。为了解决这一问题，企业开始使用一定的账号体系对用户进行设限，在各种场合提醒用户使用手机号注册或者用手机号来换取更多的权益，本质上就是希望能够在多屏时代把用户“认”出来。

网络是一个自由开放的社会，很多用户可能会出于不同的目的去注册多个账号，如果简单从账户的角度去看的话，可能会被认为是多个用户；但是，从现实的角度来看的话，他实实在在又是同一个人。在一般的场景下，有多个账号可能不会是什么特别的问题，但是一旦涉及反欺诈等需要识别到“人”的场景时，则必须要将之识别出来。所以，在这种场景下，识别用户的时候要对串联数据做灵活处理，比如一些隐秘的识别和串联数据——密码。密码不能让企业完全认定两个用户是同一个人，但是，当企业怀疑这两个账号是同一个人的话——他们的密码往往很类似或一样，由此可以判定这两个账号肯定是同一个人，为了方便很少人会给自己不同的账号设置不同的密码。当然，作为用户的你大可以放心，因为在大公司中你的密码都是被加密过的，没有人知道你真正的秘密是什么，对比密码的方式也仅仅只能是对加密后的一串字符做比较。

在美国，另一项非常重要的数据——车牌号，也对用户的识别具有重大意义。据说，美国FBI对人和车的跟踪是分开的，他们认为车牌号很重要，和人有着非常高的相关性，能串联起来用户的很多信息。

如果你想知道日常生活中哪些是很有价值的识别和串联数据，那么可以回想一下你的银行卡丢失后，你打电话到银行卡中心时对方会问你的问题。一般来说，对方会问你“你哪天发工资”、“你家里的固定电话号码是什么”等类似问题，而这一系列问题就是在把你的个人数据做一个识别和串联。因为在银行怀疑某个人是不是你的时候，生日、固定电话号码是有权重的。有可能在有了2～3个这样的数据后，即使你没有密码，银行还是会相信你，为你重新办卡。

我在运营数据的时候，经常观察这样的数据，有时候能够直接识别，有时候需要间接识别。电话号码、生日这些都是非常重要的数据，因此，我认为有些人把自己的生日放在SNS（社交网络）上，是非常不妥的做法。

所以，千万不要小看识别数据的价值，我的经验告诉我，能够辨别关系和身份的数据是最重要的。这些数据应该是有多少存多少，永远不要放弃。在大数据时代，越能够还原用户真实身份和真实行为的数据，就越能够让企业在大数据竞争中保持战略优势。

数据价值**2**：描述价值

再用T恤的例子来分析。你搜索T恤，搜索的是什么品牌、材质、尺码等类似的数据，而这些都是描述数据，可以用来刻画研究对象。研究对象可以是商品，可以是企业，可以是用户，而反过来，描述数据也可以帮助我们更好地理解研究对象。

在网络上，我们经常会听到很多关于“好男友”标准的段子，比如“身高170～180厘米、体重60～75公斤、月收入5 000～10 000元、不抽烟等”，这其实就是将“好男友”这样一个感性的指标数据化了，这里用到的数据就充当了描述研究对象的作

用。

在通常情况下，描述数据是以一种标签的形式存在的，它们是通过初步加工的一些数据，这也是数据从业者在日常生活中做的最为基础的工作。一家公司一年的营业收入、利润、净资产等数据都是描述性的数据。在电商平台类企业日常经营的状况下，描述业务的数据就包括成交额、成交用户数、网站的流量、网站详情页的流量、成交的卖家数等，我们就可以通过数据对业务的描述来观察交易活动是否正常。

但是，对于企业来说，数据的描述价值与业务目标的实现并不呈正比关系，也就是说，描述数据不是越多越好，而是应该收集和业务紧密相关的数据。比如一家兼有PC平台和无线平台业务的电子商务公司，在PC上可能更多地是关注成交额，而在无线平台上更多关注的应该是活跃用户数。

描述数据对具体的业务人员来说，能够使其更好地了解业务发展的状况，让他们对日常业务有更加清楚的认知；对于管理层来说，经常关注业务数据也能够让他对企业发展有更好的了解，以做出明智的决策。

用来描述数据最好的一种方式就是分析数据的框架，在复杂的数据中抽象出核心的点，让使用者能够在极短的时间里看到经营状况，同样，又能够让使用者看到更多他想看的细节数据。分析数据的框架是对一个数据分析师的基本要求——基于对数据的理解，对数据进行分类和有逻辑的展示。通常，一般优秀的数据分析师都具备非常好的数据框架分析能力。

数据价值**3**：时间价值

如果你不是第一次在电商网站上买东西，你曾经的历史购买行为，就会呈现出时间价值。这些数据已经不仅仅是在描述T恤了，还展示出在这一时间轴上你曾经买过什么，以便让网

站对你将要买什么做出最佳预测。

在考虑了时间的维度之后，数据会产生更大的价值。对于时间的分析，在数据分析中是一个非常重要，但往往也比较有难度的部分。我们可以通过以下两个案例来做更好的分析。

数据分析实例

> 第一个案例是电商A首页的焦点图，是一个关于茶叶的广告，但这个广告并不是对每一个人都显示，对我显示的原因是我之前在电商A上买过茶叶。这个数据的价值就是广告系统通过对我历史数据的判断，将我识别成一个可能会买茶叶的人，并通过广告系统的内在竞价机制，将这一款茶叶推荐给了我。
>
> 第二个案例是关于一个类目的演进路线，其中存在一些很有意思的现象。比如烘焙用品会有一个明显的演进路线，黄油、模具、芝士、包装盒等用品会随着烘焙产品周边产品的演进而不断呈现出新的需求。这就是在大数据的基础上，基于用户的时间来对产品的演进进行分析，以让消费者更加快捷地找到所需要的商品。

大数据一个非常重要的作用就是，能够基于大量历史数据进行分析，而时间则是代表历史的一个必然维度。数据的时间价值是大数据运用最直接的体现，通过对时间的分析，能够很好地归纳出一个用户对于一种场景的偏好。而知道了用户的偏好，企业对用户做出的商品推荐也就能够更加精准。

时间价值除了体现历史的数据之外，还有一个价值是“即时”——互联网广告领域的实时竞价（RTB，Real Time Bidding），它是基于即时的一种运用。实时竞价就是当用户进入某一个场景之后，各家需求方平台（DSP，Demand Site Platform）就会来进行竞价，对用户现实场景进行数据推送。比如，用户正在浏览一个和化妆品有关的页面或者正在商场逛街，在这个场景中就会出现和化妆品有关的信息。这个化妆品

的广告不是预先设置好的，而是在这个具体的场景中通过实时竞价出现的。

数据价值**4**：预测价值

数据的预测价值分成两个部分。第一个部分是对于某一个单品进行预测，比如在电子商务中，凡是能够产生数据，能够用于推荐的，就都会产生预测价值。比如，推荐系统推荐了一款T恤，它有多大的可能性被点击，这就是预测价值。预测价值本身没有什么价值，它只是在估计这个商品是有价值的，所以预测数据可以让你对未来可能出现的情况做好准备。推荐系统估计今天会有10个用户来买这件T恤，这就是预测。再问一些追加问题：“你有多大的信心今天能卖出10件T恤？”你说有98%的可能性，那么这就是对未来的预判及准确度的预估。

互联网的很多场景都会对单品的预测进行运用——在电商A中，哪些商品能主推，哪些商品不能主推，都是可以通过数据的预测价值来体现的。又比如电商A的团购活动，每天有大量的卖家和商品想要报名上团购，那到底该如何进行选择呢？在这个过程中就要通过这个卖家和商品的历史数据，来预测它的某个商品是否能够卖到一定的交易额，如果能卖就能上，如果无法卖到一定的交易额，就不能上。

预测价值的第二部分就是数据对于经营状况的预测，即对公司的整体经营进行预测，并能够用预测的结论指导公司的经营策略。在今天的电商中，无线是一个重要的部门，对于新的无线业务来说，核心指标之一就是每天的活跃用户数，而且这个指标也是对无线团队进行考核的重要依据。作为无线团队的负责人，到底怎么判断现在的经营状况和目标之间存在着多大的差距呢？这就需要对数据进行预测。通过预测，将活跃用户分成新增和留存两个指标，进而分析对目标的贡献度分别是多少，并分别对两个指标制定出相应的产品策略，然后分解目

标，进行日常监控。这种类型的数据能够对公司整体的经营策略产生非常大的影响。

数据价值5：产出数据的价值

从数据的价值来说，很多数据本身并没有特别的含义，但是在几个数据组合在一起或者对部分数据进行整合之后就产生了新的价值。

在电商中，这样的场景很常见。比如，在电子商务开始初期，很多人都关注诚信问题，那么如何才能评价诚信呢？于是就产生了两个衍生指标，一个是好评率，一个是累积好评数。这两个指标，就是目前在电商平台的页面上经常看到的卖家的好评率和星钻级别，用户能够基于此了解这个卖家的历史经营状况和诚信状况。

但是，仅以这两个指标来对卖家进行评价，会显得略微有些单薄，因为它们无法很精确地衡量出卖家的服务水平。于是，又衍生出更多的指标，比如与描述相符、物流速度等，这些指标最终变成了一个新的指标叫作店铺评分系统（DSR），可以用之来综合评价这个卖家的服务水平。

当然，某个单一的商品在电商网站上可能会出现几千条评价，而评价中又是用户站在自己的立场描述的，但是推及到某个用户上，每次买一样东西都要阅读几千条评价显然是不太可能的，因此就需要把这些评价进行重新的定位，以产生出新的能够帮助用户做出明智购买决策的数据，这些数据就是关键概念的抽取。我们可以看看下面的图5—1，就知道怎样让评价的数据来产生新的数据了。



图 5—1　电子商务网站中可以对用户购买行为产生影响的几大关键概念

在认识了数据的分类和明确了数据的价值后，我们就能更好地识别出哪些是我们想要的核心数据，就能够更好地发挥数据的作用。精细的数据分类，严格的数据生产加工过程，将让我们在使用数据时受益匪浅。

BIG DATA
The Business Revolution
数据化思考

用傻瓜的视角去观察

有一次，我在广东看舞狮子，一个智力有缺陷的人跑来问我：“老师，中国有狮子吗？”是啊，中国并没有狮子。中国有老虎，没有狮子。在日常生活中，我们对石狮子和舞狮子已经司空见惯，所以并没有人会去考虑这样的问题。那么，中国是什么时候开始舞狮子和门口放石狮子的呢？于是，我到处去查证。

事实上，换一种视角去观察，我们会避免理所当然的盲点。这让我想起微信的产品经理张小龙，他曾经说在做微信的时候，自己有一个秘密武器——他可以一下子把自己变成傻瓜，从傻瓜的角度去研究产品的设计方法。我们每一个人，即使是成天在嘴上说自己不聪明的人，多少会有些自恋，也会认为自己是聪明的。其实，许多惯性思维会让我们变成真正的傻瓜，因为你永远不会去思考“中国有没有狮子”这个问题，也不可能会设计一个自己才懂得如何使用的产品。

比如，我给你一张白纸，命题的关键词是“贝壳”，你会很快速地把与命题相关的东西填满。这是因为从小老师就教我们背书，背得越快越好，谁快谁聪明——在这里，快速填满相关的东西就是聪明。

如果我们从思考这个角度来讲，这种做法虽然看起来聪明，但其实很笨。因为从反向的角度来看，你填满得越快、越熟练，你或许会越快成为了一个只会复制的机器，最终失去自主思考的能力。

有意思的是，如果我们用傻瓜的视角观察问题时，我们又应该用何种视角去做判断呢？此时，我的建议是不要以短期利益为重，眼光要放长远，利及众生。

一个为了当前利益而与身边人斤斤计较的人，长此以往如何会有真正的朋友呢？这是我们经常会经历的事情，虽然视野的宽窄影响着帮助我们做出判断的信息，但决策时也不应该只以短期利益作为考虑标准。

像傻瓜一样去观察，用利及他人的眼光去判断，用一句话总结来说就是："大智若愚。先是若愚，后是大智。"

在此，我举一个例子，我觉得"若愚"的意思就是：我知道"我不知道"。记得我的前老板，有一个习惯，他在提问前总会说："我很笨，刚刚我没有听懂，你能够再重复一遍吗？"现实中，把自己放在一个很低的位置去提问，这是很聪明的做法，但这对于很多人来说其实很难做到，特别是在拥有一定的地位、一定的能力后，他们更不愿意去承认"我不知道"。但你要知道，只有你知道"我不知道"，才能理所当然地说："对不起，我不懂，你能再重复一遍吗？"

在我所在公司的高层里，还有一位非常聪明的人。有一次开会，某个同事在讲他的规划，说到一半，我们的CEO就打断了他说："等等，你说的是什么？是说我们的服务器不够用，还是说有很多服务器不能用？"总之，后面会有一堆后续的问题等你解答，这让很多人觉得我们的CEO是一个令人"讨厌的人"。但他坚持的理念是，让问题在很清晰的思路下去做执行，这是依托在非常仔细的观察和解答完所有疑问后才做决定的"傻瓜习惯"。

另外，当我们看得长远以后，许多东西就公平了许多。同样的道理，在人生里，有大爱。学会帮助别人，你会发现，整个世界的时间仿佛都被拉长了，很多东西都变得公平了，而选择就变得容易了。

06 从用数据到养数据

BIG DATA 众说《决战大数据》
The Business Revolution

吴 海 桔子水晶酒店集团创始人

“大数据”和绝大多数“新概念”一样，已经成为许多闭门造车者宣扬的主题，而品觉—— 一个被我视为数据科学家的朋友， 将他在eBay和阿里巴巴近十几年的实战经验进行梳理和总结，使“大数据”不再是一个人云亦云的概念，而是一个可以用心领会和使用的科学方法。

过去，有一些问题一直困扰着我：“现在的企业获取数据如此容易，数据的增长速度如此之快，那么对于企业来说，到底要收集什么样的数据？收集多少数据？收集数据的边界在哪里？”后来，我在美国遇到一位高人，他认为，过去收集数据很难，而现在获取数据资源变得越来越容易，但是如果收集数据的出发点不是为了解决问题，那么收集再多的数据也没有什么意义。

同时，许多企业还有一个疑问：“现在收集数据不难，成本也不高，为什么不先收集了数据再说呢？等以后需要数据来解决问题时，再拿出来用不是也可以吗？”这位高人同样也给出了这个问题的答案，他对此持否定观点，并指出用这样的理念来设计数据应用注定会失败。

数据收集是没有边界的，我为此也痛苦了好一段日子，比如，收集一个人的生日，虽然可以精确到几分几秒，但这么精确的数据又能用在什么领域，又能产生什么价值呢?

事实上，数据是有生命周期的。比如，某网站的母婴频道在主动收集用户的宝宝信息，包含宝宝生日、性别、小名、身高、体重，这里我们就要清楚这几个数据的用途和生命周期，如宝宝的身高和体重。图6—1就为我们展示了这几大内容。但是，妈妈们填写的是宝宝当前的数据，我们知道宝宝的身高、体重变化非常之快，也就是说这两个数据的生命周期很短，在当时的推荐中可能有用，但过了一段时间后，这两个数据就失效（Data Broken）了；而宝宝生日和性别这两个数据的生命周期就很长，可以从生日的年份中推算出宝宝的年龄，而性别则基本是终生稳定的。



图6—1 某电商母婴频道宝宝档案创建页

保存数据及其收集时的背景（Context），也是一件不容易的事情。还是以收集宝宝的年龄作为例子来阐述，我们可以通过用户购买特定的商品集合，如几段奶粉和尿不湿的型号等，

知道宝宝所处的大概年龄段，以推算出宝宝当前的年龄。但如果仅仅是保存宝宝几岁这个数据，那么这个数据很快也会失效，因为人的年龄是在动态变化的。同时，你还需要保存宝宝几岁这个数据获得的时间，即在哪年哪月计算得知这个宝宝的年龄，而这个信息就是背景数据。另一种做法更加聪明一些，即通过宝宝现在的年龄反过来推算宝宝是什么时候出生的。

所以说，在收集数据时，我们必须知道这些数据未来可以用来做什么，如果今天都想象不出来的话，日后就更不可能了。

打一个比方，如今很多电商高管会询问数据分析师商品的重复购买率是多少。于是，工作人员收集数据来计算重复购买率，却很少想到高管们需要重复购买率是来做什么决定的，这就好比“刻舟求剑”的故事。事实上，在世事多变的大数据时代，我们不能只是机械地套用方法或指标。

重复购买率有不同的定义，而做不同的决策需要考虑经过不同定义的重复购买率。如果一家投资公司想收购A公司，那么会从重复购买率来看整个A公司的运营良好程度或用户质量等；如果从A公司本身运营的角度来看重复购买率，那么它更应该关注的是日、周级别的重复购买率的变化趋势，或者当月新增客户有多少人在三个月后进行了重复购买，从而可以衡量出每个月新增及存量客户的忠诚度和质量，进而找出改善的空间。在知道了以上的背景之后，再去选择使用什么数据不是更有的放矢吗?

数据应用因小而美

“小”不是指数据量，而是指应用的目标很具体。

2011年年底，作为一名数据分析师，我开始思考怎么从“用数据”转变为“养数据”（即从数据化运营转变为运营数

据），这段时间我为收集什么样的数据而特别烦恼。而且，我也曾经试图做出一个特别大且适合多数人使用的数据应用，可是后来发现这在数据应用的起步阶段几乎是不可能的—— 一是找到可以解决大部分人需求的数据应用并不容易；二是当时公司的数据非常丰富，需要考虑的因素很多，因素之间的联系又很复杂。

所以，当开发数据应用的时候，数据就等于原材料；当原材料一直处于变化的情况下，做出来的产品就很容易出问题。体会到数据和应用的关系之后，我最后决定从小角度切入，先把小应用做出来，这就是很好的瞄准器。

这里说的“小”指的是应用的目标很具体。请注意，“小”不是指数据量。许多人在没有获取足够的数据，并且缺乏对数据理解的情况下做出决策，其实是在“享受”自己的无知。打个比方来说，对于一款数据应用，如果我的目的是分辨两种决策谁更好以及差异在哪里，这就是一个很具体的问题；但如果我的目标是想知道如何让公司赢利，这就是一个空泛的目标。

经过这样的一番周折后，按照小角度切入的想法设计数据应用，就可以做得具体而快速，而且可以避免因原材料的变化而导致数据缺失的问题。

把数据放进“框”之中

在大数据的背景下，必须考虑数据之间的关联性。一个单独的数据是没有意义的，要把数据放在一个“数据框架”（场景）之中，才能看出存在的问题。

为了把这个问题说得更明白，这里我拿前阵子一家电商公司找我讨论的问题来举例。

数据分析实例

A公司问我，要不要撤去导航网站的广告？因为它怀疑许多老客户是从导航网站访问官网，而不是直接访问官网。把这个问题说得再直白点，就是他们想弄清楚在导航做广告与公司业务之间的关系。那么，接下来就要观察用什么“数据框架”有助于其做出决策。

1.公司当前的投入产出比

- 明确导航网站引进的新老用户占比如何？
- 引入的新老用户的投入产出比和转化率如何？
- 推断撤去导航网站，对流失新老用户的影响如何？

2.与竞争对手的博弈

有一个问题可能会被忽略，那就是你不做导航网站广告，你的对手就会立马投放此广告。做数据框架的时候，要特别注意框架不是静态的，而是存在博弈的，需要把竞争对手的因素考虑进来。

3.考虑时间因素

建立框架时要考虑时间因素。

- 用现在、过去和未来的眼光来审视导航网站，看看导航的质量是不是越来越好；
- 需要注意的是时间有延迟性，引进来的流量会有一些延迟，在两三个月后才能知道新用户的价值。

总之，“数据框架”是商业分析师的灵魂所在，可以让其从框架中找寻问题的关键因素及答案。不同的问题有不同的框架，不能完全在此进行阐述。

如何用框架来做决策

想要解决的问题越复杂，框架也就越复杂。但是，决策最重要的前提是要从小角度切入，从“小”做起。

对此，我总结了4步走的方法：

1.首先确定有什么问题，从解决问题的角度出发去收集数据。

2.把收集到的数据整理好，放入一个“数据框架”内（这个框架是用来帮助决策者做决定的）。让决策者用框架更清楚地看到数据与决策之间的关系，比如A公司在框架内要知道竞争情况、新老客户的比例情况等因素以及多种因素之间的关系。

3.看框架与做决策的关系。比如，A公司与导航网站有三种选择——完全不合作、部分合作和全面合作。数据分析师就可以根据数据框架告知A公司该怎么决策。如果发现数据框架与决策不能匹配，就必须返回到第2步。

4.根据决策行动，然后检查行动是否达到目的。如果行动后发现根本没有达到目的，就要检讨整个链条，寻找问题出在哪里。是数据有问题吗？还是因为框架不对？或者是决策不对？是否还有数据没考虑进去？

所以，想要解决的问题越复杂，框架也就越复杂。对于现在多数还没有开始做数据应用的电商公司来说，框架一开始千万不要过于复杂，一定是要针对某个需要解决的具体问题开始搭建框架，令框架与决策紧密相关。所以，在搭建框架前问问自己：“你的问题是什么？你的决策是什么？反过来你的框架又该怎样搭建？”虽然你的回答可能多种多样，但是，最重要的前提是要从小角度切入，从“小”做起。

养数据，重要的数据战略

> “用数据”更多的是一种方法论，而“养数据”则是一种数据战略，是基于深入业务理解的更高层次的商业决策。

“养数据”还有一个重要的含义，就是要决定收集哪些数据。这个决定不容忽视，因为这可能是公司数据战略中非常重要的一个环节。因为，很多公司的做法，是有什么数据就收集什么数据，完全没有主动性，但这样做遇到的一个核心问题就是，公司自主思考的能力开始缺失，对该收集什么样的数据不甚了解。事实上，在我看来，被动收集数据的行为是“收集”，而主动收集数据的行为则是“养数据”。再拿上面讲到的某电商母婴频道的例子来说，如果收集主体是一家医院，那么宝宝的姓名、性别、出生年月和妈妈的姓名等数据的收集，就是收集数据，因为在住院档案和出生证明上，这些都是必不可少的。但是，如果是母婴用品要针对家庭进行营销，那么这些信息显然是不够的，宝宝奶粉的品牌、家庭收入、职业等信息则需要企业自己主动去收集，而这些数据也就是“养”出来的数据。

养数据通常有两类，一类是网站自身没有的数据，需要用户主动提供的；另一类是公司拥有的，但没有进行收集的数据。

若要收集公司没有的数据，在“养数据”时通常需要花费更多的精力和技巧。举例来说，现在国内外有很多个性化的手机购物应用，他们会根据用户的偏好进行商品的推荐，让每个人都有完全不一样的购物体验。具体的运作原理是，当用户第一次使用这些应用时，界面中会弹出一些问题，通过互动的方式来收集用户的信息数据，比如会出几款衣服，让用户挑选哪一款是更喜欢的。这样，在几个问题之后，应用就知道用户的喜好了。这类应用通常还会有另外一个特点，就是会让用户点“赞”，这个点赞的功能，就是让用户明确地告诉应用自己喜欢什么，这样用户就可以在过程中，不断地告诉应用自己的偏好。这样一来，应用就可以 “养”出用户的核心数据。

对于公司有数据但没有收集的状况，则更多的是因为内部资源协调的问题和公司决策的问题。很多提供餐饮和团购类信息服务的网站，通常会有一个功能是“把信息通过短信发送到手机”，这样就不需要用户进行特别的记录，不过有些网站会记录用户的手机号，而有些网站则不会。那么，这些手机号需不需要企业存下来呢？这就是一个与业务决定和资源协调相关的问题，如果决定是记录，则必然会需要开发人员在后台进行功能的优化，如果不记录，那就会丢失这部分信息。我的观点当然是记录这些数据，也即是“养”数据，因为这不仅能够校验用户的手机号码是否有效，还能够慢慢地建立起企业与用户的社交关系。

总之，**“**用数据**”**更多的是一种方法论，而**“**养数据**”**则是一种数据战略，是基于深入业务理解的更高层次的商业决策，**,**数据养的时间越早，积累的数据也就越多。养数据同样也是一种管理和商业艺术，在养之前可能谁也不确定最终会出现什么后果，但一旦养成，则会产生非常大的商业价值。

BIG DATA
The Business Revolution
数据化思考

远离“或”选择

我们做的很多决策都是“Yes”或“No”，比如看电影，看或不看只有一个选择。

在遇到压力的时候，我们非常容易走进“或”选择的阶段，会一直盯着“去”或“不去”看电影这个选择，而不会考虑“可不可以去喝茶”，这也是选择，但在我们心中却是余地非常小的选择。因此，当我们面对这样两难的问题时，一定要从压抑中迅速跳出来思考：可不可以不是“或”，而是“也”。要不要去看电影？不去或者也去。

比如，有一天，你走进一家古董店，看到了一件很喜欢的东西，就问：“老板，这个多少钱？”老板看到你进来时眼睛一直盯着这个古董，于是聪明地对你说：“您是今天的第一位客人，很有缘啊，您随便开个价吧。”于是，你开价500元。老板马上说：“不行啊，这个价值1 500元。”你说：“不，我就出500元。”老板说：“好吧，今天第一笔生意，当送你了。”事实上，500元成交的这件商品或许只值100元。在这个过程中，其实只要你一张口开价，便意味着老板已经成功地把你引入“或”的选择题中。老板心里笑了：“吃定你了。”开价，就等于在给你一个“或”式选择：500元，要或不要？没有其他。

但是，如果你足够聪明，在看到老板想要把你引入“或”的选择时，应该立即跳出来。此时，你无须开价，而是问：“多少钱？”老板说：“1 000元。”你说：“为什么值1 000元？”老板会说：“这个是明朝的东西。”你再问：“怎么证明是明朝的啊？”这样之后，你会觉得豁然开朗许多。

别把自己逼到“或”式选择的境地，这在逻辑思考中非常致命。你必须跳出来，问：“为什么是‘或’？”

讨价还价的时候，停一停：“为什么是1 000元？”

讨论今晚是否看电影时，等一等：“为什么要看电影？”

别人说数据增长20%时，想一想：“为什么是20%？”

如果你要做一名数据分析师，你脑海里就要长期装着这样一个想法：“这个数据是怎么来的？”你要打破局限，不要把自己封闭在狭隘的思想中。在未来商业中，谁被逼到“或”式选择境地，往往谁吃亏。

“或”式选择在电商的活动中也很常见，“促销的最后一天”就是给你一种压力，让你不知不觉走进“买或不买”的选择中。那么，这时你该怎么做呢？你应该快速地退后一步思考，为什么是最后一天？其实，这是一个假信息，其实前两天就已经是其所谓的最后一天了。就像在街头，我们常见的那种店——最后三天促销，但三天又三天过去了，半年过去了，还是最后三天大促销。这些店就是在骗那些刚好今天走过、以前没有来过的新顾客，把他们推到“买或不买”的境地。其实，对这些新顾客来说，他们不走到这里，就会还有很多选择。

如果你足够细心的话，会发现小孩大都用的是“或”式选择。小孩要是问现在能不能看电影？如果你否决了，他不会建议去打高尔夫球。在没接受教育之前，小孩都不会有这样的想法。在经过教育之后，他们才慢慢开始懂得怎么做出选择。但长大之后，在外界的压力下，很多人往往又会变回小孩，逼得自己没有选择的余地。

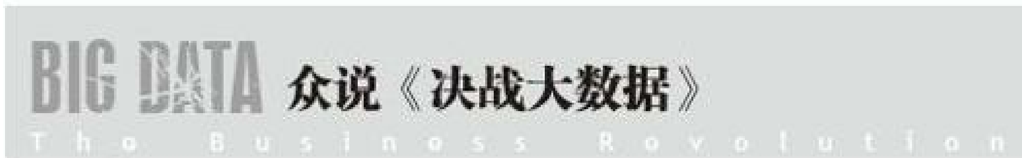
如果把这样的思考应用在我们的成长之中，很多人就是受“或”的概念影响，将自己不成功的原因归结为某件事情没有做好，他们不会想到失败的结果可能是由于自己没做过另外一件事情。

我在澳大利亚曾经勤奋地练了两年的击剑，击败了很多人。但我发现，当我上了新台阶后遇上高手时总是输。教练对我说，如果要摆脱现状，就要加强防守。教练说：“过去你赢对手，是因为进攻很厉害，而同级中比你差的选手会因为你的快速进攻而败下阵来。但现在你想快速进攻，就会发现胜利的可能性已经很小了。事实上，这时候不是你的

进攻不够快，而是防守不够好。当你防守很好的时候，就可以做到假进攻真防守，或假防守真进攻，因此当你攻击时，对手不会太早防御你。你发现没有，别人输给你，是因为怕防守，而不是怕进攻。”

同样的道理，在面对“或”式选择的时候，我们更容易被惯性牵绊。很多时候，你需要远离“或”的概念，才不会被困住而浑然不觉。

07 数据的盲点，负面数据的力量



张溪梦 **LinkedIn**商务分析部总监

品觉的书主要讲的是决战！那么咱们就从兵法的角度来看大数据！只讲三点：

第一，对“道”的理解：就是分析师对人和事物基本规律的诠释。分析问题千万不要从分析大数据开始，而是要从对人、世界、产品或者商业行为最基本的认知着手！

第二，对“计”的理解：计就是计谋！交战之前用“商业智能”的一个重要作用就是要造成信息情报不对等，然后进一步造成了战略优势的不对称，从而造就取得优胜的“势态”。

第三，对“胜”的理解：决战的目的是要胜利，兵法上取胜的一些基本的要领比如以快打慢、以少胜多、以众击寡，在大数据分析上完全适用，要做到分析得快速而精准、大规模部署以及产品化等。

这次为品觉的新书做推荐，既感到非常感激和荣幸，又感到诚惶诚恐。我的水平非常有限，完全是抛砖引玉，希望读者们细细品味《决战大数据》！

如果数据存在盲点，核心数据就无法轻易显现出来。那么，分析有盲点的数据就像是在清水中炼油，虽然耗时耗力，但仍有可能会一无所获。那么，为什么有人能在相同的数据中找出关键点，有人却视若无睹？从大数据时代出发，在思考的过程中，有一个问题必须得到解决，那就是假定数据都是可以获得

的情况下，什么样的数据才是关键？

数据盲点

在数据中，盲点可以分为两类：一类是物理盲点，另一类是逻辑盲点。

既然谈到了盲点，就必须要对盲点下一个定义。我们通常意义上的盲点指的是，眼睛没有看到的现实存在体。数据上的定义也基本类似，就是指被我们无意中忽略了的重要数据或者角度。在数据中，盲点可以分为两类：一类是物理盲点，另一类是逻辑盲点。

所谓的物理盲点，就是指在数据库中不存在这样的数据，即企业没有收集到应该收集的数据，这一类数据问题的产生通常是数据收集策略出了问题。让我们来回顾一下在商品详情页上作标记的这个案例，之前我们在评价某一个商品是不是被用户有效地浏览时，使用的数据指标是用户在页面上的停留时间，但是这个指标存在天然的缺陷。

首先，停留时间不完全代表用户对商品的喜好。在企业收集到的数据场景中，有很大一部分可能是用户的无意识停留行为，比如，用户刚打开网页，就被领导叫出去开会或谈话等，而这种情况导致用户在某些页面停留时间过长的数据就没有收集的必要。其次，停留时间通常情况下不容易计算清楚。在传统的日志模式下，停留时间都是按照下一个页面打开的时间作为上一个页面结束的时间。当一个人在多个窗口同时停留时，则会陷入停留时间过短的误区。在要评价用户是否对页面感兴趣时，数据就出现了物理盲点——企业没有精确的数据来描述用户是否对商品感兴趣。面对这种情况，就需要企业在商品详情页上用技术的手段作上标记，以记录用户在页面上是否有动作以及是否滚动屏幕去看更多的内容。在明确了用户具体看到了哪些内容后，企业就收集到了相应有效的数据，就

能够很好地对商品页面进行后续的分析。

对于物理盲点来说，如果出现在PC的网页上，算是“万幸”；如果出现在手机客户端，那就是一个比较大的问题了。为什么这么说呢？因为PC上网页的模式是网站可以随时调整的——上一秒做完改版，加上代码发布后，下一秒用户就能看到全新的内容，企业需要的数据也会在下一秒开始收集。但是，物理盲点一旦出现在无线APP上，就会给数据收集带来很大的难度。

造成这种现象的主要原因有两个：第一个原因是无线终端的数据收集技术还不是很成熟，即便是已经想到了要收集手机的哪些数据，也会由于开发工程师的疏忽等非主观原因造成数据的忽略，当然，PC端也可能会有疏忽。而第二个原因会让这个问题加重，在PC端，如果出现了因疏忽造成数据遗漏的情况，企业可以通过快速地修复网页来弥补，一般只会造成几分钟或者几小时数据的遗漏。而无线APP客户端就没有那么容易了，因为无线终端要求用户经常更新版本，但实际生活中，并不是所有用户都会安装新版本。另外，即便是所有用户都愿意安装新版本，新版本的推出也会受到系统方的制约，比如，对于iOS系统而言，申请发布新版本可能需要等待数日。这时候，整体的数据收集活动就会遭受很大的影响。所以，在数据收集的规划期，一定要尽可能地想清楚自身需要什么数据来衡量业务，尽可能地避免可能出现的盲点。有时候，为了避免盲点，企业必须尽可能多地收集一些数据，以此来减少物理盲点出现的可能性。

相对于物理盲点的数据遗漏情况，逻辑盲点就是有数据但是没有被很好地发掘出来。数据逻辑盲点的出现很多时候与数据分析师或者数据使用者的经验和敏感度都有关系。在分析数据时，相关从业者需要对数据抱有敬畏之心，不可轻易放过任何一个可能产生问题的点。

通常，企业在分析数据时，会将焦点放在宏观的数据上，比如总体登录商品页面浏览的人数。很少有人会去关注个别商品页面的浏览人数，而且即使是有些商品的浏览人数明显激增时，分析师往往也不会投入精力去详细研究——由于日常工作的繁忙，这类数据更多时候是被当作一个异常数据忽略了，而这就形成了一个盲点。逻辑盲点如果未得到重视就很有可能对企业的业务造成极大的损害。

数据分析实例

> 曾经，在某电商网站就出现过这样的一个案例——某一个商品的浏览人数非常多，而这个商品的交易量却很有限，这与常理是不相符的，因为通常情况下，大量的浏览一定会带来大量的成交，那么，为什么会出现如此违背逻辑的现象呢？后来，该电商网站的一名员工发现，他虽然没有浏览过这个商品，但却在自己的浏览记录里发现了这个商品，奇怪之余，他马上就上报了系统故障问题。直到这时，大家才投入精力去重点解决这个问题。后来，经过认真的分析发现，是有人用了一种非常“聪明”的方法在刷数据。该电商网站的用户ID是一串有序增长的数字，而此商品的卖家就模拟其他用户的ID去访问它商品的所在网页，并且成功骗过了日志系统的防作弊机制，使得没有浏览过这个商品的用户在自己的浏览记录里面发现了这个商品。那么，这个卖家为什么要做这样的事情呢？因为网站会在后台记录这个产品，并且误认为它是用户的喜好，并向用户推荐相关的产品。

在上面这个案例中，从问题出现到问题解决经过了一段时间，虽然没有造成重大损失，但也对部分用户的体验造成了影响。其实，这样的问题如果能够在出现之初就得到关注，并认真地去研究异常数据、寻找根源，就能够避免其带来的影响。

同样，还有一些数据也容易被忽略掉。对于一些平台型电商或者自营电商的平台来说，成交额总是能够获得更多的关注，而对退货情况则关注得较少。在各家电商中，通常又都会

对成交额定指标，每个运营类目的团队或者员工都会有成交额的KPI。在很多时候，考核成交额并没有太大的问题，但是就是有些公司去钻“重成交，轻退款”的漏洞。这种情况在各个公司中屡见不鲜，一些经营平台的员工会让平台的卖家“想方设法”去完成成交额，并许诺将活动资源的支持作为回报。所以，有些卖家就采取了“自己买进，然后退款”的手段把成交额提上去。这样一来，虽然员工得到了业绩，卖家得到了资源，但是公司却没有获得任何好处。其实，在这个过程中，只要关注类目或者商品的退款率，这些数据就能够被轻易地发现，盲点自然就会消失。

在逻辑盲点中，最大的盲点是将**PC**数据和无线终端数据混着看。很多网站都会统计转化率的数据，即电商网站通常所看的购买用户除以浏览用户得出的数据，在单纯PC的情况下，这个转化率大体可用，还是可以描述宏观经营状况的。但是，在无线终端进入后，这个数据就有问题了。无线成交的数据记录在了成交中，但是浏览的数据却没有计入到浏览中，所以就会导致转化率越来越高。如果忽略了这一点，就有可能漏过重要的商机。

数据除了物理和逻辑盲点，甚至还有一些是人为制造的“盲点”。人为制造的盲点就是故意把数据进行掩盖，或者人为地调整数据的口径。在一些网站，定义独立访客（UV，即UniqueVistor）的时候考虑更多的是自主UV，也就是说只有用户主动访问一个页面时才算数。但是，有些人为了自己的绩效，可能会做一些手脚，比如，在对外投放弹窗广告时也算用户主动访问；在页面加上嵌入模块，使得用户访问一个页面时有两条数据以增加UV的数据。这些情况通常比较隐蔽，甚至能够欺骗管理层。所以，管理层在听取数据汇报的时候，不仅要看数据的变化和趋势，也需要对数据的口径有一定的理解，最重要的是要问清楚定义是什么。如果涉及转化率，就要问清楚计算它的分子分母是什么，这才不至于掉入数据的盲区。幸

好，现在商业的整体环境对数据都比较重视，我接触到的高管对数据也越来越关心，越来越多地使用数据做决策，也越来越重视数据的分析和挖掘对业务产生的价值。这样一来，数据的盲区也会越来越少，数据也会变得越来越有价值。

小偷思维

> 面对数据的盲点，正能量思考告诉你怎么做可以到达成功的终点，负能量思考则告诉你怎么做才不会失败。

数据的盲点有没有什么价值呢？让我们举一个生活中的例子。

“十一黄金周”时很多人都外出旅游，你的房子里7天都没有开灯。这样一项信息，对你而言没有任何用途，但对于小偷来说就是机会。小偷的成功，另外一个很重要的原因在于能够度量风险。对于小偷来说，一间屋子三天不亮灯并不代表没人在家。小偷会一直观察环境有没有突然变化，会对零散的信息如噪音特别敏感，他将所有感官能力都提升到了警戒线。有些小偷之所以会成功，一个很大的原因在于他们会度量风险，也就是会使用“负能量思考”（Negative Thinking）看待问题。

人类是有惰性的，我们喜欢尽量把自己想要思考的东西变得简单。“坏人”却很愿意承认不可预见性的存在。就像我们常常判断一件事情是“好”或“不好”，而不是“未知”。举例来说，我参与了一场赌博，赌注为200元。在这一事件中，以我看来，赢就是“好”，输就是“不好”。但对于“坏人”来说，他们就一直在衡量一些“未知”的情况，比如，你去的赌场是不是黑店、警察会不会突然闯入没收赌资或者刚一开始你就输了200元。

“坏人”之所以会对“未知”的情况如此关注，是因为当“坏人”的风险实在是太高了。对于正常人来说，家里丢失了一件

东西也许无关紧要，而对小偷来说，失败是要坐牢的，这就注定他要细心观察那些正常人所看不见的数据。想做小偷里的“成功人士”，首先想的肯定不是如何偷东西，而是偷完以后怎么不被人发现，小偷必须懂得利用蛛丝马迹来降低风险。

其实，当失败的代价越大、风险越高时，人类自然有这种天生的动物般的敏感。也许你不知道自己已经在风口浪尖，但当你已经身处其中时，你一定会提高自己对风险的度量能力，不是吗？动物世界就有这样的现象，能力越小的动物，越具有优秀的聆听风险的能力，而猛兽们则没有。

也许你会问，对于我们这些不准备做“小偷”的人来说，负能量思考又有多大的意义？

我们会受到经历和经验的影响，如果自己真退化到很弱小的状态，就会像小动物一样能够聆听风险。高风险能让一个人对不可预见性非常敏感，那时候的负能量思考和避免风险的能力就自然而然地出现了。所以，如果我们平时也懂得用这样的“思维”思考，我们就会避免很多错误。如果这时我们再站在“坏人”的角度去做负能量思考，觉得别人出错就是你的机会，很多人出错就是你的大机会，这甚至会产生一条很有价值的产业链。

在美国和澳大利亚，有种灯可以设置偶尔亮、定时亮和晚间亮，这说明它的研发者是一个典型的具有负能量思考能力的人。可见，我们身边的这些负能量思考，是可以产生巨大价值的。

再看看我们身边的商业环境，这样的负能量思考也很常见。奇虎360的周鸿祎就是具有这样思考能力的人。看他的轨迹，从3721到360，他注意到了很多人的错误，观察到了很多人需要免费杀毒软件的需求。

当你希望能达成某个目标时，你必须用心去观察别人观察的事物，这样才会提升你的观察力。对一般人来说，都意识到了培养自己“正能量思考”（Positive Thinking）能力的重要性，也都意识到要用心观察成功人士是如何成功的。但同时也要注意，应该培养自己运用负能量思考的眼睛去观察别人是如何失败的。面对数据的盲点，正能量思考告诉你怎么做可以到达成功的终点，负能量思考则告诉你怎么做才不会失败。

当对手的数据就摆在那里时，不看的人才是傻瓜，这并不涉及道德问题。偷看不是数据世界的本质，但观察别人怎么失败却可以实实在在地降低自己的成本。值得注意的是，获取数据情报在正常情况下很不容易。在这个时候，你得想着不能破坏公司的名誉，还要想着不能让别人来偷看自己的数据，更要把别人来“偷”数据的后门给关掉。

在上述过程中，我只是在“模仿小偷”。总的来说，一个“坏人”观察到的数据，正常人很难看得懂，但如果你学会用负能量思考的方式去观察，必然会对你有很大的提升。如果你想要训练到这样的程度，是一件很难的事情，毕竟你的心态是正常的，你不会去关注谁的包没有拉好，不是吗？无意识地培养负能量思考，就叫乱想。如果你为了一个目的去规范训练的话，你就会成为一个很厉害的“小偷”；对于日常生活来说，叫作很厉害的“观察者”；而对于数据世界来说，你就是一个很厉害的“数据分析师”。

“小偷思维”给了我们换个角度观察数据的方式，而现实中我们可以用这种思考方式来度量数据盲点的价值。这两个内容看似分离，但是核心点是一样的，就是有没有看到应该看到的数据，有没有错失不应该错失的数据。

数据盲点的价值

> 是否看到数据盲点价值的核心就是，有没有看到应该看到的数据，有没有错失不应该错失的数据。

先提一个假设性问题：200米以外忽然发生了爆炸，你如何通过大数据鉴别出是不是有人蓄意而为？如果是，如何快速地锁定嫌疑人？

通常情况下，我们可以主要分析两个数据：在爆炸之前，现场有没有出没一些形迹可疑的人；爆炸发生后，人群中有没有行为怪异的人。打个比方说，一般爆炸发生后，大部分人都是朝着事故发生的方向走，以观察谁受伤或者单纯为凑热闹。如果这时，有两个人是反向快速离开现场的，就可以算是怪异行为了。

有人也许会问，这和大数据有关系吗？当然有，而且还能引发我们思考另外一个问题：数据的盲点在哪里？

平时大家接触到的都是信息，而其中可以被量化的信息称为数据。所谓大数据时代，就是指可以记录和调用的数据变得越来越多。互联网和移动互联网就像两台巨大的超级传感器，把人、物、空间和时间等数据进行实时记录，让我们在时间和空间里都留下线索。

数据从太少到太多，从简单到复杂，而且还衍生出了无限的关联数据。面对海量数据，我们常常疑惑的是，这些数据究竟是不是足够了？是多了还是少了？回到前面提到的案例中，人们常常可以预见充满“正能量”的信息（现场围观群众的行为），但很容易漏掉充满“负能量”的数据（把数据范围放大，你可以注意到急速离开现场的人）。这里面的盲点指的是，我们观察数据的时候容易只将焦点放在正面数据上，而忽略掉负面数据。

这种案例不胜枚举。以我的日常工作为例，在团队给高管

们做报告的时候，我通常不看报告，而是主要记录高管们“什么时间点头”、“什么时间针对性地做了笔记”和“什么时间没怎么听在看手机”。你不要小看这三个数据，当与当时报告的内容对应（关联）起来时，我就能容易地知道在刚才做的报告中哪一部分是不吸引人的和哪一部分是需要改进的。尤其是高管看手机的时间，完全可以算是理论上的“负能量”数据，但对报告的改进其实非常有价值。

再比如说，公司每月有10亿元的交易量，虽然额度已经很大了，但你是否分析过，其实还有2亿元可能的交易量没有做成。为什么？因为有人看了不买。为什么不买？通过分析这些“负能量”数据，其实很可能发掘出非常有价值的平台交易增长点。

还有一类盲点数据，常常发生在我们赖以依存的经验上。我的导师说过：“‘假定’这件事情，一般是所有误判的源头。‘假定’会让人盲目。”很多时候，人们在做一个决策时会依靠经验，但经验有可能是错的，这会导致“假定”的错误，继而影响执行的结果。如果今天你看到一张桌子，上面有很好吃的食物，但把手伸过去时却被烫到了。以后再看到类似的桌子，你恐怕都会因为这个之前的经验而不敢靠近。显然，错误的经验会让你不断地犯错误。

经验会成为盲点，还有一个重要的原因是你没有注意到环境变化的趋势。有时候，有经验反而会经历惨败，因为这件事情本身已经出现趋势的变动，但大家却未察觉。之前的经验关联到当下的决策，会变成你认为必然发生的经验，甚至是规律，但这个规律会骗人。股票市场中的教训已经太多了。

做数据分析，经常会被邀请做趋势判断。比如，对两家公司的员工增减态势做分析。从表面上看，一家公司的员工比另一家少，但如果再看行业的背景，一个是“新创公司”，一个是

有10年历史的公司。此时，再结合业态的发展趋势时，你就会对他们做出不同的判断。想要做出正确的判断，必须先抓住其发展的趋势。

古语有云："无知最穷。"在大数据时代，那些自诩对大数据什么都懂的人，其实却对无数盲点而不自知。他们浅薄地认为数据越多就越能解决问题，实则是无知的表现。

在海量的大数据面前，我们面对的最大盲点就是，衡量应该收集多少数据才算足够，即存；如何管理从未见过的如此多和广的数据，即管；如何使用它们实现更大的价值，即用。

为什么数据会骗人：常态、时态与变态

有一部讲述魔术的美国电影《致命魔术》，片头一开始就对观众发出了一句考问："你真的看见了吗？"很多人以为自己看见了，但因为缺乏真正的观察，其实还是获取不到真相。

什么是"观察"呢？观察就是在事物内部找到其相对性，而且是从共通性里面找出相对性。所以，当我们想要判断一个人讲述的事实是不是谎言时，就要基于以下两种情况进行判断：一种是要看，这个对你说话的人，是你人生当中第一次见而且见面频率很低；另外一种，则是在日常生活中，这个人你重复见过很多次。这对于判定一个人是否说谎具有不一样的意义。因为一个熟悉的人，判断其是不是说谎其实很容易，而一个你只见过一次面的人，判断其是否说谎，则要更难一些。因为此时你没有经验对比，只能通过短时间内的观察，这种相对性很难提取。

在此，我们首先要了解两个词语：常态和时态。在说常态之前，我们先了解一个概念"情态"——蕴藏在体内为"情"，表露在体外为"态"。情有七种，《礼记·礼运篇》中指出七情为：喜、怒、哀、惧、爱、恶、欲。欢喜、愤怒、哀伤、恐惧、爱慕、厌恶、欲望等，都是人们内心的情态。这七种情态在人体内无间断地扰攘，被人的边缘系统激发之后，会衍生出时态。有些人赌赢了，高兴到旁若无人的境地，其实就是边缘系统出了问题。当他冷静下来也许会惊讶于刚才的表现："呀，刚才失态了，真丢人。"

我们在观察一个人的时候，最容易观察到他的时态。但如果要观察出一个人时态的差异，就必须了解这个人的常态。常态是什么呢？常

态分为以下6种：弱、狂、哗、周旋、慵懒、媚。所谓弱态，是指言语轻细、动作温柔、包容性强，像小鸟依人一般。所谓狂态，则是指不修边幅、倔强好胜、言谈举止旁若无人。我们可以对比自己身边的朋友，就可以轻易判断出他的常态。一个属于狂态的人在被问及意见时，通常愿意主动地表达观点，并坚信自己是对的。一个属于弱态的人则不会表露太多意见，会给出模棱两可的意见。哗态的人，通常遇到事情会大手一挥："你们都别说，让我来说。"周旋态的人被问及选择时，说到A会说A很好，说到B会说B其实也不错。慵懒态则比较漫不经心，什么都不是很在乎。媚态则意指谄媚。

在人们的日常状态中，每个人并非单独属于一种态，而是经常混在一起。但这里有一则交友贴士，要远离弱而且媚的人，如果是狂而且媚的人，其实是非常厉害的人。

在我们日常做判断时，一定要先知道这个人的常态，然后他才有判断的依据。当你向一个常态是慵懒态的人汇报工作时，他忽然呈现出了正襟危坐的时态，你就可以试着判断，这次工作的内容也许特别重要。

要想区分谎言，要先区分常态和时态。而如何知晓一个人的常态呢？就要靠观察。所以我在之前提到，对熟悉的人更容易判断其是否说谎，因为了解其常态；而对刚见面的人则比较难做出相应的判断。

在我们的日常用语中，还有变态一词，其实就指的是由常态变为时态而衍生的异常肢体动作或是不理智行为。如果洞悉了这两态，其实就很容易观察出一个人"变"态时的起点和变化点。

另外，还有很重要的一点要观察。当一个人发生常态的改变时，你就要思考促使他改变的是什么环境和事件？比如说，当一个人被警察抓到警察局时，这个人会恐惧，会暴躁不安，肢体动作会粗鲁，但这不能推导出其一定是恶人，因为环境的压力也应当考虑进去。但如果是一起爆炸案，在爆炸发生前后，是否有不符合常态的人出现在现场中呢？案发前在现场是否有不正常停留的人？案发后当有人受伤倒地或围观人群过来救护时，有两个镇静的人行色匆匆拿着手机反方向离开现场，那么这两个人的犯罪嫌疑就很大，因为这就属于常态中的"变态"。

以人推"数"，当我们在观察数据时要考虑以下几个问题：数据的常

态和时态是什么？为什么数据会骗人？这些通过直观感受和仔细观察之后得出的结论也许会大相径庭。

第二部分

阿里巴巴的大数据秘密

数据化运营实践过程中的问题


日常报表信息量大，难以捕捉有效信息
管理层：那么多报表到底我该看哪些？
日报
周报
月报
仪表盘
竞争性报告


信息分散在不同部门，缺少有效整合
管理层：今天出现业务异动了吗？


业务异动的处理往往是自上而下来推动
负责人：理论上今天应该没有银行的接口在进行升级，我马上回去分析问题到底出在哪里。
管理者：为什么今天信用卡网关的成功率只有50%？是哪里出现了问题？


关键分析成果取得实效，但未实现沉淀
发现问题
分析问题
发现根本原因
设计解决方案
部署验证
选项1
结束
选项2
业务分析逻辑固化
自动业务监控
我们现在按选项2针对每一个异动分析，业务逻辑实现固化了吗？

08 阿里巴巴的大数据实践

按照我们通常的理解，一个人在注册某一网站的时候，性别登记只会是男性或女性，但不知你是否能想到，阿里巴巴竟然有18个性别标签。

在经过仔细调查后，我们发现了一些可以识别用户性别的方法。比如，某个登记用户早上的浏览行为更男性一些，晚上就会变得更女性（也许是妻子在用）。真实的性别只有0与1的关系，而现实却是0~1的关系，或者70%是男性，30%是女性。在每一个场景内，性别的变化都确有其原因，例如搜索、社交和广告等。静态的“真实”性别在A/B测试[1]中的表现不如动态的虚拟性别有效，这恰恰证明了数据是需要运营的，绝不可以闭门造车，这样的例子在阿里巴巴不胜枚举。

谈到数据化运营，在公司内部我们是这样的流程：例如，我们要为一个童装类目做营销推广以征集新客户，我们会先找到目标客户群，会把对这个类目感兴趣的消费者用标签找出来，然后通过发邮件或者短信等方式来吸引他们的关注，这其实就是数据化运营，也就是用数据去帮助企业解决问题。最简单的解决办法是找出所有曾经看过但没有购买的客户群，又或者利用关联类目去扩大目标客群。

而数据从“用”上升到“养”，即运营数据，我们就会尝试在整个淘宝用户中查找。以针对12岁年龄之下儿童的商品为例，此时我们的做法就是在用户中搜索有多少用户家中有12岁以下的孩子，但却未发生过从这个类目购买商品的行为。你会发现，以前只能运营有过购买行为的几百万用户，现在我们的运营范围竟然可以达到几千万可能有12岁以下孩子的用户。从几百万到几千万，这就是运营数据，它是一个需要从主动收集数据到运营数据，再到产生新数据的过程。

在我的理解中，从数据化运营到运营数据是不断运行的循环。在这样的循环中，会容纳许多新的、不同维度的数据，这些数据经过在整个循环中的适应过程，然后再运用到数据化运营中，并且改变原有的运营方式，这个过程就是我认为的大数据落地的方法。

数据化运营的前提是假定数据是稳定的，并且以此来改变企业的运营。但以前的数据都是结构化数据，比如说，早期大部分企业都用财务数据来帮助企业运营，因为财务数据经过多年的积淀，相对来说较为稳定。而自从非结构性数据或者半结构性数据产生以后，数据化运营的难度就变大了。比如，音频是一种典型的非结构化数据，你很难把音频变成结构化数据，有了这样的数据以后，用数据来指导企业运营，面临的技术难度远远大于以往。

我们做数据化运营，其前提是假定数据是稳定的；而运营数据，则是假定数据都是可以获取的，而且是不稳定的（见图8—1）。

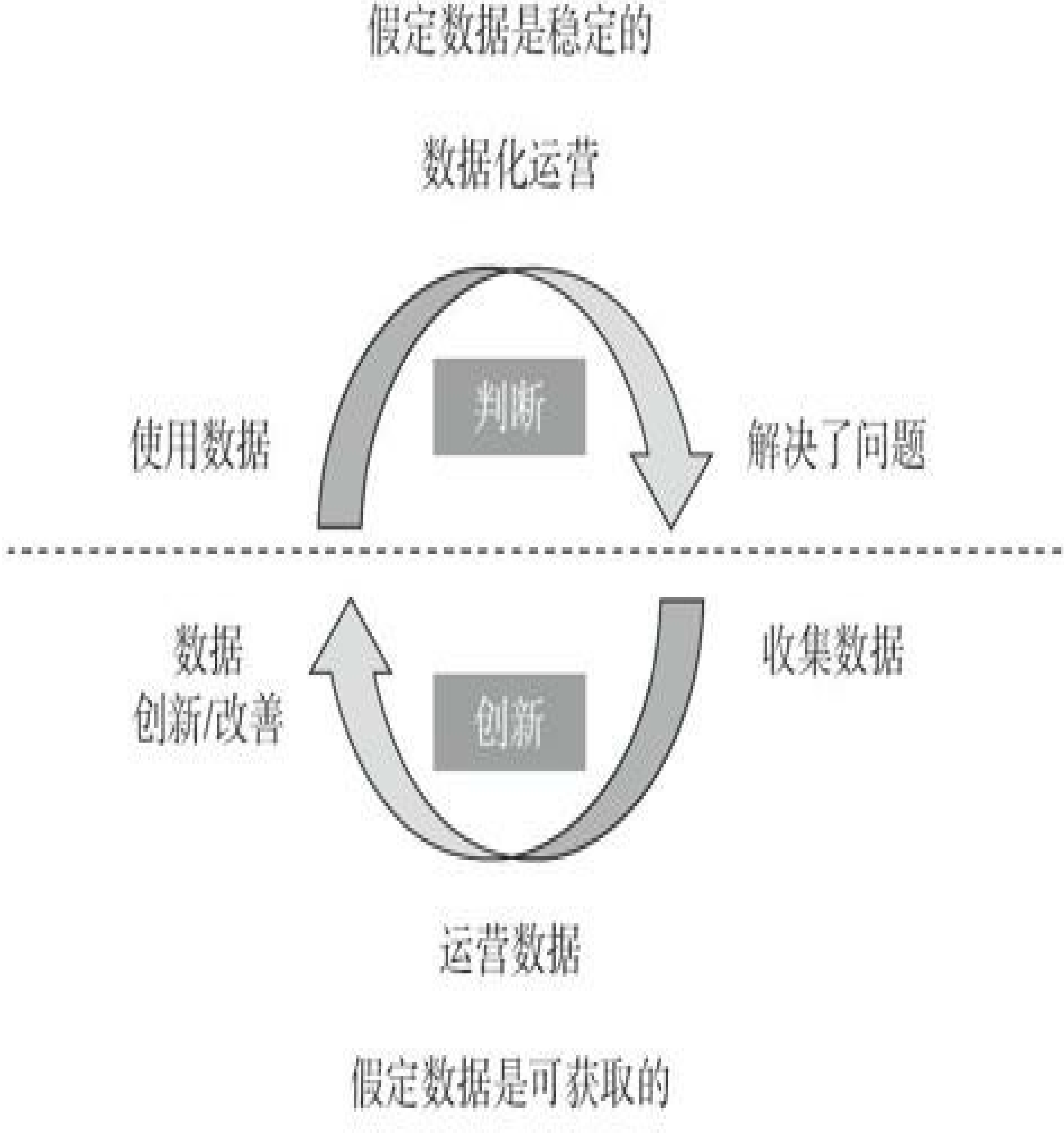


图 8—1　数据化运营和运营数据的闭环系统

这又如何理解呢？所谓稳定，比如说关键词ROI，这项数据听起来很稳定，但反映的却是短线利益。数据化运营需要和商业咬合得非常

紧密，所以数据也是混合在商业里，以假定稳定的方法去做业务上的对比和细分及趋势预估的。

这并不是说稳定是对的，而是因为企业对数据运营的分析是需要假设很多稳定因素的。

假定数据是稳定的

假定数据是稳定的，也意味着我们习惯于不去寻找一些新数据。

在这里，我们用点击率做例子。数据运营人员可能会觉得这项数据我们一直在使用，能有什么特别的呢？但在我眼里，点击率是变化的。我的问题是，当用户点击到详情页时，又有多少用户连第二屏都不会拉下去看？如果这项比例很高，那么就意味着点击率本身就有很多水分。

所以说，在数据化运营里，我们往往假定数据稳定不变，因为在我们的认知中，我只是数据的使用者，我假定数据是稳定的就可以了，这是普遍心理。那么，为什么我要去关注数据稳不稳定？而且，假定数据稳定，也意味着我们习惯于不去寻找一些新数据。

我的钥匙不见了，一般的做法是去那些我可能丢东西的地方找。同理，在运用数据的时候若出了问题，我们也会是一样的心理，不会去思考除惯性思维之外的其他可能性，而且这些地方无法解决问题时，我们就会不解。而这是数据化运营中很常见的一条死路。

于是，到了养数据，也就是运营数据的阶段时，我们的做法就必须改变了。此时，当我们要想静态的性别时，就要问自己：“会不会不好用？是否需要改善？”

从企业的角度来讲，数据化运营一般是自上而下的——管理者强调KPI指标，员工们按照目标进行分解。所以，数据化运营是计划性的，但凡计划性的一定都是稳定且结构化的。

运营数据却是自下而上的，经过了无数的错误和无数的实验，慢慢地向上追溯。就像阿里巴巴有18个性别这件事情，这18个性别不是管理层凭空想象出来的，而是将众人发现的性别经过交叉比对，分析到底

在什么场景下更适合。可见，这样的结果是试出来的。

在我们的日常工作中，你很少会发现负责业务的管理者说："我要收集这项数据，它很有用。"这样的场景一般不会出现。我有过这样的经历，是因为我是数据部门的管理者，我这么做的原因，是在跟其他公司的人讨论时产生了这样的想法，而我也会在我回到公司后提出"我们也试试这个数据"的意见。比如，前面提到的那个亚马逊根据IP地址观察用户附近10公里内有没有书店的例子，同样的运营数据的方法我就在淘宝尝试过。

我想说的是，如果没有了运营数据的这个部分，收集来的数据绝对就不是大数据。

假定数据是可获取的

"用数据拿数据"的方法可以将数据化运营和运营数据打通。

从假定数据可以获取的角度来思考问题，是数据化运营里很重要的一个方法，这与"如何利用数据来拿到更多数据"的问题并不一样。今天，我已经知道了你的一些数据了，但是并不多，此时我就要想办法用我知道的东西来拿到你更多的数据，这在数据化思考里叫作"有意识地用数据拿数据"（Data Get Data）。

现在，最常见的一种"数据拿数据"的方法就是"A/B　Test"　。原理是，我本来知道你对某种东西很感兴趣，今天我给你做一个测试，本来是A页面，我给你一个B页面，然后看看你有什么样的反应（有点像功夫里的假动作）。其实，B页面拿到的数据肯定是基于A页面拿到的数据的。

比如说，你经常会走过一条路，而今天在你走过这条路时，我安排了一个美女从这里走过，然后测试一下你的路径有没有变化以及你有没有停下来。

正常来说，以前收集到的数据会告诉我你在路过的时候是不作停留的。而今天你停了下来，那我就得到了"美女的出现可以让你停下来"的数据。这时候，美女就是我抛出来的一个问题，即用问题来获得更多数据，这就是在用一部分数据拿更多的数据时采用得最常见的方

法。

将数据化运营和运营数据打通，有可能会创造出一些意想不到的创新。

我们在大数据环境里运营数据时，有可能走不通的原因是，现在的数据已经庞大到需要管理才能到创新的数据化运营。这个循环能否成型，往往需要很长时间的努力。

互联网金融就是一个近些年比较新颖的例子，现在也有很多前沿公司在做类似的事情。这么做的原因是，公司发现用户的资金链是评估其经济状况的一个有用信息，而收集到的这些数据应用在信用评估上时，可以更动态地决定用户的信用额度，以减少借出方的风险。

比如，这个网站拥有关于整个公司流水的数据，大概知道用户的资金流水——原来还有30%的订单在处理，还有不少订单在路上，而有些则是押在了一些担保交易上，所以这个用户不是没钱，而是钱押在了供应链的各个环节上。万一用户不还钱，网站也有能力卡住用户的资金流，这使得网站有能力降低自己遭受坏账的风险。

其实，阿里巴巴在从数据化运营到运营数据的循环系统中，能做到现在的成绩，跟管理层更相信数据、更愿意付出以及更愿意投资有很大的关系。此循环系统的关键在于管理层如何应对数据价值的断层，这对整个循环能否走下去的影响非常大，而阿里巴巴应对这个循环的内功，我认为答案在于“混、通、晒”。

先不要一头雾水，“混、通、晒”简单来说，应该是我们数据分析师的“内三板斧”，我会在第9章中进行详细的介绍。

[1] A/B Test，A/B测试，是数据分析中的一种对比分析方法，可以将试验组和控制组的试验结果更加明确地展示出来。——编者注

BIG DATA
The Business Revolution
数据化思考

先开枪，后瞄准

前几年，有一个关于淘宝的非常普遍的论调甚嚣尘上——淘宝正在经历着野蛮生长。

什么是野蛮生长？这个概念现在常常被我们提及，但事实上，野蛮生长应该是在资源有限的情况下才成立的，在资源丰盛的时候，或许就不存在野蛮生长的概念了。现今的淘宝，已经是一个有钱有资源的“富二代”，实际上已经不存在什么野蛮生长的前提了。

在我看来，马云是一个非常懂得利用“野蛮生长”的领导者。在一次会议中，有位同事讲解了一个提案，但是在同事说的同时，马云一直在摇头。会后，马云说：“我给你3个目标，你用一年时间去执行这件事情。”这才叫野蛮生长，因为这3个目标就是在资源有限的情况下进行的。为什么马云一直摇头，不相信这件事情能做成，却依然让这位同事去尝试呢？后来，马云提起当初他想做阿里巴巴时，也被很多人否定了。所以，即使这个同事的提案看起来成功的可能性很低，他还是要让其去尝试，因为不试怎么会知道结果呢。

对啊，不试怎么知道呢？但“试”也是有方法的。

我经常听到很多淘宝卖家说：“先开枪，后瞄准。”我常常思忖，这句话在逻辑上怎么能讲得通呢？正常的逻辑应该是“先瞄准，后开枪”啊。后来我知道了，其实这句话是有前提的，那就是：当你没有数据的时候，如何在探索之中寻找到有参考价值的数据，并且最后发现目标，做出决策。所谓开枪，就是在一堆海量数据里盲选，而瞄准则是在海量数据里盲选出出有价值的数据。这样，“先开枪，后瞄准”才是有价值的。

将这一观点进行延伸，有关开枪和瞄准的顺序问题，就像现实中，很多人容易把目标和目的混为一谈，目的和目标区别在于：“目的”是要到达的终点；而“目标”则是衡量到达终点的价值标准。

用一句话来总结野蛮生长就是：“是，但不是；不是，但是。”

这句话又怎么理解呢？我们用“是，但不是”举一个例子，我们常常会这么评价一个人，“他是一个好人，他也是一个坏人。他是好人是因为他在外面对朋友很好；他是坏人，是因

为对自己的家人很坏。”再说“不是，但是”，比如说某个人创业失败了，但是不是完全失败了呢，这其中还是有很多数据价值的。

比如，我很欣赏NALA的刘勇明，他原本是一个学海洋学的科学家，懂得如何探索和观察。后来做了商家，对用数据解决问题这一领域并不是很得心应手。他很重视客服，最擅长的做法是通过客服来观测数据，通过最前端来决定后端，这就是探索。我很赞赏他说的 “发现交易量急速上涨的时候，就要停掉广告” 。这句话体现了他作为一名科学家的思维逻辑——当交易量急速上涨时，对订单的处理能力就会下降，所以减少广告可以很好地控制局势。

“先开枪，后瞄准”的说法对于创业公司来说是正确的做法。对于创业公司来说，管理者不需要把眼光放得过于长远。如果从短期目标出发，在不断的观察之中发现问题，并且解决问题，很可能会更容易获得成功，而事实上很多人的成功就是这样造就的。

09 混、通、晒，阿里巴巴数据化运营的内三板斧

当谈到阿里巴巴的数据化运营时，我首先想到的就是“人”，我们太多的时间都在讨论我们应该要做什么，却很少会反过来想如果要让数据化运营落地首先要从“人”做起，所以在这里想跟大家分享的秘密是，数据化运营的内功是什么呢？简单来说，就是利用好“混、通、晒”这三板斧。

混，“混 ”出数据

> 只有具备商业敏感的数据分析师，才会懂得使用什么数据来驱动公司实现经营目标。数据部的人要和业务部的人经常“混”在一起。

现在的很多数据分析师，当面对专业范围内“怎么算回归”、“怎么画函数”的问题时，他们常常游刃有余，但在实际工作中却往往缺乏商业意识。如果数据分析师缺乏商业意识，那么公司就成了“盲人”——分析师不知道该使用怎样的逻辑去分析数据，而公司的决策层也得不到任何有价值的参考意见。

现在绝大多数CEO都在抱怨，每天都要看一大堆零零散散的数据。造成这种局面的原因就是，数据分析师们只是单纯地把数据传递给了管理者，却没有向管理者解释这些呈现用户行为的数据和能够在商业上产生价值的数据之间的内在关系。

CEO是没有多余精力来解读页面浏览量（PV）和独立访客（UV）等数据的。他们只需要知道数据是否有问题、反映了什么问题、最近有什么新的发现以及需要我们做出什么样的改变。

简单来说，具有商业意识的数据分析师，在监测到网站上婴儿车销量增加的情况时，基本就可以预测到奶粉的销量也会随之上升。而且，也只有具备商业敏感的数据分析师，才会懂得用什么数据来驱动公司实现经营目标。

数据分析师要怎么才能拥有商业敏感呢？我觉得就要靠“混”。

什么是“混”？我先举一个例子，比如，我以前公司的CEO就跟我说：“你每周要跟管理业务的负责人吃两顿饭，最起码两顿，这就是你的KPI。”

我认为，这就是“混”。作为数据分析师，如果不和业务部混在一起，又怎么会知道业务部在做什么。商业敏感是要靠“混”出来的，它并不会凭空出现在你的面前。从更深层次来说，数据部的人要和业务部的人经常混在一起，比如参加业务部的周会、业务规划讨论会、头脑风暴等，或者去业务部定期轮岗、互通周报邮件，一起喝茶、吃饭等。现在在公司里，我最不希望看到数据分析师长时间地坐在自己的位置上埋头苦干，我更愿意看到他们和业务人员“混”成一片。

比如，我让数据分析师们在发给我的周报里，一定要讲到业务方的动态。而且，我给他们的考评标准是，千万不要让我看见业务方发过来的周报里面有的内容，你的周报里没有。我认为，要实现这一要求最基础的出发点是，数据分析师一定要跟业务方沟通，才有可能服务于他们。

通，打“通”“混” 的数据

> “通”是“混、通、晒”里最关键的连接点。知道带着业务问题来看数据或者带着数据来看业务，这就是做到了“通”。

当你与业务人员“混”得足够熟时，在看到某些数据后，你自然就会明白，“哦，这个数据跟商业决策绝对有着莫大的关系。”

当前，各电商公司在评估公司经营状况时越来越依赖数据。但是，在今天，很少有电商敢完全肯定地说，自己掌握了体现公司状况较完整的数据。对于公司主管来说，一是因为很多电商在开始收集数据时，会发现数据非常散乱，分布在不同的数据收集渠道和运营人员——公司的核心人群手里，这就使得数据流非常“堵”；另一个问题是，绝大多数电商缺乏大数据运营的经验，只是收集了很“散”的数据，却不知道如何利用，也不知道让哪些数据关联起来。

从客观角度来看，数据运营的方方面面都可能存在影响数据精准

度的“噪音”。数据本身是客观的，但它很容易受到产品和运营人员的影响——产品目的的导向会影响运营人员的想法，而运营人员的想法则会影响样本获取的精准度，这就造成了数据在不同人眼中出现不同结果的情形。以转化率为例，市场部门和运营部门对于“转化率”的想法是不同的，如果公司内部的数据标准没有“打通”一致，公司决策时被数据迷惑和误导的可能性就会被放大。

因此你会发现，问题最后还是要归结到“人”和“公司”上来。如果不能“通”到商业环境里，即使数据很多也没有任何价值。坚持带着业务问题来观察数据或者带着数据来观察业务，兼备二者的敏感，就是做到了“通”。有些人在很短的时间里就能判断出数据是否有价值，就是因为“混通”了。

值得注意的是，“通”有两个场景。比如说，现在有一个商业场景和一堆数据，这两者产生关系时，就是商业模式和数据彼此的**“通”**，我们可以称它为**“数据中间层”**，简单来说，它就是能够敏感地反映出商业变化的数据群（**Smart Data Set**）。

另外一种更深入的**“通”**，就是存在于公司组织中的数据。譬如，有一个商业问题，但是需要数据A和B、C、D互通，才能解决。举例说明，今天的退款非常多，我会去做的第一件事情可能是去察看营销部门的数据，我看到今天UV猛增，所以退款才特别多，这是可能合理的。但是，我从客服数据中发现一些退款是少量的卖家产生出来的，那就另当别论了，因为这说明这个商业场景也有可能是作弊，此时就需要对营销、客服和风控等多组数据综合起来进行解释。所以，把数据打通，一是商业理解和数据之间的“通”，另一种就是部门数据和部门数据的交叉。

现实中的“不通”，也有两种：一种是从人的角度来说的，比如说人在思考上的“不通”；另一种则是从公司的角度来说的，比如在架构上的“不通”。这两者角度不同，造成的损失也会有差别。大部分的公司都会出现“不通”的情况，比如，部门与部门的隔离打乱了整个流程的连贯性，财务部门对产品、市场和运营等的不敏感，可能会把数据打散，这样一来，商业活动就根本没有办法继续进行。

想做到数据的积累和沉淀，想要打通数据，建立合理的系统是不

二之选。首先，做好数据安全工作，以保证公司内部不同职位的员工可以察看不同的数据；然后，统一不同部门的数据标准，使公司内部数据有统一的接口，避免混乱；最后，关联不同部门的数据，创造机会让数据的运营可以扩散至数据部门之外。

“通”是“混、通、晒”里最关键的连接点。以前，数据量没这么大的时候，公司“混”完就“晒”了，完全凭借商业敏感去运营数 据。而现在海量数据成为主流，“通”也就成为了运营数据不可或缺的一部分。

晒，“晒”出“混”和“通”的数据

> 数据能不能做到在获取、使用、分享、协同、连接、组合之上让自己变得超级简单和便捷，这是数据化运营里面非常重要的一点，这也正是“晒”的内容。

“晒”是一种在“混”和“通”的基础上产生出来的最终的数据表现，是基于人、商业和数据结合后的一种看数据和用数据的方法论。

在“晒”数据层面上，通常是通过数据来回答这几个问题：业务好还是不好，数据如何改变可以让业务更好，如何利用数据帮助业务发现机会，甚至产生出新的商业价值。这些问题看起来是递进关系，其实不然，因为具体应该用数据解决什么问题，要根据业务的场景来决定。每一个问题都有不同的回答方法，接下来具体讨论一下这些方法。

数据需要有框架可依：数据如何来证明业务好还是不好

这个问题用更专业的语言来说，就是如何搭建数据的框架，并通过数据框架来解答业务水平的好坏。所谓的框架就是，对一个业务进行指标化的分解，并通过有限多个指标来客观描述业务的状况。指标化的分解是为了能够快速地定位到问题。

举例来说，每个人一生中都可能因为感冒去医院看病，很多医生一般都会先让病人验血，再从化验单里判断是病毒感染，还是细菌感染，或者其他感染因素，然后根据化验单的指标来对症下药。而数据框架的作用正是如此。

比如，某电商网站今天的成交额提升了30%，那么这是不是一个值得高兴的数据呢？我们无法直接给出答案，此时，就应该让数据告诉我

们，假如是因为投放了外部广告，才使得成交额提升，那么，这就是一个客观的描述，我们要做的不是看成交额的高低，而是应该评价广告效果的好坏。

在电子商务领域，评价业务水平通常有两套指标，一套是常用的计算成交额的方法，即流量×转化率×客单价=成交额，它能够评价一个类目、一个商品的健康度。另外一套指标在商品大促的时候使用居多，即预热期加入购物车的商品数×商品单价×经验转化率×经验成交额占比=大促成交额，因为经验转化率和经验成交额占比通常是两个固定的数据，所以只要看看购物车有多少商品及其价值，基本上就能算出来大概会有多少成交额了。这就是数据框架的作用，它使我们分清楚了业务的状况。

但是，业务水平的好坏需要进行不断的比较，单独一个数字是没有意义的，比如180这个数据，从表面上看，没有任何意义，但是如果有比较，这个数据就有了意义。如果是身高180厘米的男性，那就算是高个子，因为中国人的平均身高可能只有170厘米；如果是汽车时速180公里，那就是危险速度，因为高速公路的限速是时速120公里，它已经属于超速50%，是危险驾驶。推及业务上，业务需要进行比较才能判断好坏，而比较的前提就是要不断地去寻找比较对象，拿电商的大促活动来说，至少需要进行三个参照组的比较才能区分业务水平的好坏，分别是：这次大促对比平时的涨幅、这次大促和上次同等规模大促的涨幅、这次大促同比涨幅对比上次大促同比涨幅，经过这些比较，才能客观地评价大促的效果到底好不好，如果没有选对参照系，那很有可能得出错误的结论。而要选择正确的参考系，就需要分析师具有很强的商业敏感，而这种商业敏感就来自 “混”和“通”。

除了搭建数据框架之外，好的展现形式也必不可少，最典型的就是图和表。但是，现实中很多人不知道图和表到底该用在什么地方，现在就介绍几个小诀窍。如果业务场景是要查找精确的数据时，用图，就错了；如果业务场景要观察变化趋势却用了表，那也错了。一个好的分析师，不论是设计数据产品还是做图表都需要有很好的“晒”的意识，而且还要让“晒”出去的数据能够吸引到更多人的注意。

数据如何改变可以让业务更好

这一过程就是通过数据发现问题，并解决问题。就如之前说到的化验单的例子一样，通过数据的分析找出问题的症结，并进行有针对性的优化，这样的例子在阿里巴巴几乎是天天都在发生，以下就是一个相关的例子。

数据分析实例

淘宝有一个“二手闲置”业务，它与淘宝主站笔记本类目有一个合作项目——以旧换新。用户在二手闲置频道“甩旧”（卖出二手商品）后，就可以获得在笔记本类目以优惠价“买新”的资格，而且在“甩旧+买新”两步骤都完成后，还可以获得数百元的集分宝。从项目描述来看，整个交易原理并不复杂。

那么，在项目上线后，整个市场反馈如何，用户是不是买账呢？根据淘宝后台的数据显示，整个活动的流量漏斗效应非常严重。大量参与活动的用户中，最后成功“以旧换新”的寥寥无几，整体转化率非常低。那么，问题出在哪里呢？事实上，这还是要从数据切入查找原因。

- “估价”到“售旧”环节用户流失率最高；
- 在流量来源监控中，其中一个渠道的转化率几乎为0；
- 手机及平板的回收转化率明显高于笔记本用户；
- 在已成功回收的商品中，宏、联想、三星等品牌占比最高；
- 在成功购新的商品中，价位以3 000~4 000元为主；
- 在来访用户中，男性居多。

通过上面这些数据反映出的问题，我们可以做以下4大方面的改进。

- “以旧换新”整体步骤过多，漏斗效应非常明显——需要首先在“估价”到“售旧”中优化并简化流程；
- 尽管“以旧换新”给活动带来了不少流量，但这些流量几乎无后续转化——流量来源需要调整；
- 用户对品牌属性、价格高低是有偏好的——橱窗展示商品的品牌和价格也可优化；
- 来访用户中男性居多，细化分析后发现少男（20~25岁）和轻熟男（25~30岁）为主要群体——广告定投。

基于上面的数据和发现，后续的数据落地动作也就清晰起来，具体做法有如下3个方面。

- 流程以简单化为原则，对于用户流失严重的环节，进行较大调整。

- 流量精准度。“以旧换新”项目二期流量投放上选择以少男和轻熟男为主的流量源，放弃有流量无转化的渠道。

- 商品选择。将联想、宏等品牌放在较明显的位置；推荐商品价位下调（由原来的7 000元以上下调至3 000~4 000元或4 000以上为主）。

经过这样的调整后，淘宝“以旧换新”二期活动的效果明显，整体转化率提升了6倍左右。

如何利用数据帮助业务发现机会

利用数据帮助业务发现机会就涉及具体数据在业务场景中的应用了，这里最核心的价值点就是使数据变成人人都要用，而且是人人都必须用。我曾经提出过一个理念，就是人人都是数据分析师，也就是说让公司的业务人员都能够通过数据去做决策，让数据来驱动业务，以实现真正的数据化运营。比如，有一个团购网站就做得非常好。这个网站的运营领导和一线运营主管都有非常好的数据意识，他们和数据分析师“混”和“通”了之后，非常清楚地知道数据能够解决什么问题，并和数据分析师一起对网站的主要流程进行了剖析，最终得出了以下5个观点。

1. 网站流量和页面有限，必须个性化；

2. 商品的进入是公司成败的关键，必须选对商品；

3. 销售过程中可能会出现断货的情况，必须有销量预测；

4. 价格是竞争的核心，必须关注竞争对手；

5. 服务是核心，必须关注发货速度和用户投诉。

定位了这些问题之后，就开始着手进行解决，对于个性化、选品和销量预测等问题，数据分析师迅速跟进，并进行数据模型建设；对于价格和服务的问题，他们建立起了专门的预警系统来进行观测。这样一来，数据运营得非常好，对业务的提升也起到了很大的作用。

用数据如何创造出新的商业价值

为了说明此问题，我在这里举一个阿里金融的例子。阿里有一句话，叫作“让信用变成财富”，阿里金融也正是这样的。阿里金融的核心就是通过数据来计算客户的信用水平，并通过信用来对客户进行授信，

让客户能够方便快捷地获得所需要的资金。基于这种信用模式，越来越多的相关产品被开发出来。其核心就是通过业务和数据的结合，将数据“晒”出来，从而形成一种竞争力。

“晒”是产生力量的过程，以上是数据化运营中的“晒”，同样适用于企业内部。对数据分析师来说，说服别人才是“晒”，如果从产品场景来讲，就需要说服产品经理把某个东西放在产品里面，这就是数据工具体验上的一大难题了。公司架构上没问题了，但是数据工具不便于使用，用户体验自然就不令用户满意了。

所以，数据能不能做到在获取、使用、分享、协同、连接、组合之上让自己变得超级简单和便捷，这是数据化运营里面非常重要的一点，这也正是“晒”的内容。数据分析师一定要让数据变得简单，只有这样才不会出现前面所说的——收集数据的不知道怎么用，用数据的不知道数据是怎么来的。也正是因为数据很复杂，假如数据分析师无法将之变得简单，管理和运营的工作人员自然会产生困惑，也就不可能真正做到“通”和“晒”。

阿里巴巴的“混、通、晒”三板斧其实是配合了数据方法论与人的修练，能做到借事修人， 让用数据的人在数据中成长，循序渐进地让每个人都成为数据分析师。

思考，要学会关窗口

我一直觉得，每个人每天的思考能力和池中的水一样是有限的。人和人不一样，有些人可能思考能力强一些，就可以做出更多的思考；有些人思考能力弱一些，则只能做更少的思考。但不论思考能力强弱，每做一次思考都会让自己一天的思考池中的水流减少，从而降低自己的思考和决策能力。

有这种想法的不仅仅是我一个人。我曾经在《哈佛商业评论》网站博客中看到过一篇文章，很有意思，题目好像叫作“Boring is Productivity”，中文直译过来就是“乏味，是一种生产力”。文章里面提到了一个案例，有人问美国总统奥巴马：“为什么每天你穿的都是一样的衣服？”而奥巴马的答案是：“因为这样可以减少我的思考时间，我就可以将更多的时间留下来做更有价值的决策了。”这样的例子在文章中可以找到不少，其核心意思就是每天我的思考能力是有限的，如果我要做更加有意义的思考，就不要把有限的思考资源浪费在不能产生价值的思考上面。

这一观点，套用社会上一个比较时髦的词语来说，就是思考也要注意“节能环保”，我们要保护的是自己有限的思考资源。对于如何做到节能环保的方法，我自己使用的是“关窗口”的方法。我们每一个用电脑的人都清楚，如果在电脑中打开过多窗口的话，会有两个可能的影响：一个是使电脑的负担增大，电脑的运行速度变慢；另一个是即便电脑的性能足够强大而不至于让窗口影响它的运算速度，但我们的大脑也无法同时处理多个信息。我们能做的就是把多余的“窗口”关掉，认真地去考虑一件我们认为非常重要的事情。

“关窗口”在计算机世界中很容易执行，但在现实生活中却很有难度，尤其是在动态的工作过程中。为什么我会觉得困难？因为我觉得在我们思考的过程中，永远不可能只有一件事情可以做，对于管理者来说尤其如此。不断地被其他人影响，不断地被即时聊天工具的弹出窗口打断，不断地被电子邮件提醒去做事，这一切都会让我们的思考过程中断。所以，才会有不少时间管理方面的学者建议，每天处理很多事情的“忙人”可以在一个固定的时间段里去处理电子邮件，以提高办事效率。

但是，一旦到了动态的过程中，典型的状态就是在一场谈话或者一场会议中，那就不像不查收电子邮件和不开即时聊天工具那么简单了。谈话的过程是动态的，而且这个过程可能不会严格按照会议主题的规定来进行，而会被各种细枝末节的小问题和各种情绪所干扰，最终导致可能浪费了不少时间，却没有收到应有的效果。对于这样动态的谈话过程，我自己的做法是

永远只开一扇“窗”。这个意思是说，在我的谈话过程中，我的大脑中始终会有几个问题被放在首位，这几个问题不会随着谈话的进程而随意改变，这些问题包括：我这次要解决的是什么问题？对方（包括客户、同事等）要的是什么？现在提出的方案解决了这个问题吗？在只开一扇“窗”的前提下，如果会议中出现了不符合主题的内容，我一般不会过多参与，或者有意识地去结束这些与会议无关的话题。

但在之前，我参加一些电商的会议时，总是会被一些细节所干扰，即便最终提出了建议和解决方案，也可能不是真正的解决方案。比如，在会议之前，我可能想通过解决产品的问题来提高转化率，不过一些问题一旦被放到会议中讨论，就可能会受到各种因素的影响——管理者可能会说是产品的问题，而产品部门总是有数据来说明他们自己的产品没有问题，实质是运营能力有问题，而运营部门又会拿出种种数据来说明运营的动作是有效的，是吸引到的用户有问题。结果往往是会议离主题越来越远，如果再碰到一个不是特别善于演讲和辩论的市场部负责人的话，问题的矛头最终可能将指向市场部，市场部就会背起这个“黑锅”，至于到底是什么问题引发了转化率低，其实谁也不知道。

这个时候，对我们干扰最大的就是数据，每一份数据一定都是经过了精心准备的，看起来在逻辑上又是非常严密的，对于那些快速思考的人来说，就非常容易获得认同。所以，此时对我来讲，就一定不要被各种其他因素干扰，而在每一个细节上的过度争论都会减少我对最终核心问题的判断
精力。

我在心里时刻会告诉自己，我的精力是有限的，不要把时间浪费在不值得的地方。这又是一句说起来很容易，但是做起来非常难的事情，比如说，在开会的时候，并不是每个人都是决策者，他们一定会从自己的角度出发，一定会提出各种观点，此时我会做的事情就是闭上眼睛，好好地想一想，我们到底要解决的是什么问题，我应该用什么逻辑来解决，在抛开一切杂念获得验证的思路之后，再来看细节问题决策就会显得非常简单了。其实，用逻辑来思考的问题或者发现的问题，就是一个本质问题。

10 存、管、用，阿里巴巴运营数据的外三板斧

从2011年开始，阿里巴巴开始从数据化运营向运营数据发展，并形成了一个良性循环——从数据化运营的内三板斧“混、通、晒”走到了运营数据的外三板斧“存、管、用”。

存，数据收集的开始

收集数据不是目的，让收集起来的数据如何产生价值才是最终的目的。

“存、管、用”从存开始，第一件事情就是把数据收回来、集中起来、存起来。

为什么要收集数据？因为在以往的数据化运营里产生了一些问题，虽然我们解决了一些，但还有一些问题并没有得到解决。当谈到如何更好地解决这些问题时，我们想到了运用数据可能是有效的，所以在这个场景里面，我们开始收集数据、管理数据和使用数据。

就拿电子商务的数据收集为例，比如母婴类目，我们可以通过填写宝宝的年龄或者购买的特定品类（不同段数的奶粉、不同型号的尿不湿和不同尺码的衣服）等来推算出宝宝最新的一个年龄阶段；在汽车类目上，我们可以通过客户购买的机油、滤清器等型号来推算出客户是否有汽车以及车型；甚至是衣服也可以用来作为数据收集使用的场景，比如从一个用户购买衣服的历史尺码来观察用户是否有身材上的变化。关于数据收集，前文已有较多篇幅的介绍，这里不再赘述。

所以，就数据的收集而言，最重要的不是看我们收集了什么数据，而是要思考这些数据如何使用以及收集这些数据到底能够起到什么样的作用。用一句话来说，就是收集数据不是目的，收集起来的数据如何产生价值才是最终的目标。不过，如何收集在未来具有价值的数据的确是一个难题，当中就需要一些经验的判断了。

管，保护好存储数据

学会用数据产品来解决获取及使用数据的问题。

数据存储下来之后，数量和广度都很大，就需要对之进行完善的管理。数据管理的内容包括很多方面，比如，数据的来源、如何让数据不丢失、如何保护数据的安全、如何让数据准确和稳定以及如何更好地运用数据，这些都是数据运营中的“管”。但是，“管”并没有一个标准可循。在我和一些投资人以及国外互联网公司的数据专家们的交流中，他们都认同我的一个观点：大数据管理到底要怎么做？答案就是“没有准确答案”。

其实，对于数据的管理，整个大数据行业和其他行业一样都经历过很多起起落落。比如，IT技术刚兴起时，各公司对IT系统和OA的追随十分盲目，最后却没有产生期望内的回报。到后来，信息化管理开始兴起，管理信息系统（MIS）开始被很多企业推崇，不管规模大小，企业都开始使用ERP和MRP系统，结果，仍然没有达到预期。就数据而言，在互联网公司，2004年左右，美国的一些数据管理经验在国内造成了很大的轰动，很多公司纷纷建立BI团队，但是到了2009年左右，各公司又开始不完全认同BI数据部门，这也是事实。

但也正是在那个时候，国内顶尖互联网公司的数据化运营开始启动。阿里巴巴的数据管理最不一样的地方在于非常依赖数据产品，希望用数据产品来解决获取及使用数据的问题。

当时，我们犯了一个直到现在才发觉的错误。那就是，我们的做法跟一些美国公司的想法一致，即“不管怎么样，我们先收集数据，将来肯定有用”。

我觉得，在我职业生涯的这些年，最值得反思的就是这句话。因为没有一家公司可以让你无止境地收集数据，然后再使用，这根本是不现实的。

而这就是“不做决定的代价”。因为，在这个世界上，有一些决定是我们一定要做的。从运营数据的角度来说，如果我们只收集数据而不做分析和应用的话，代价就是很沉重的存储成本。

无须多言，这种存储成本的代价是巨大的。即便是一家富有的公司，即便是它的机器比较多，但也只能短时间地延续这种损失。即使是

阿里巴巴这样的公司到最后也会发现，这样做是不行的。因为不管你有多少机器，这些数据都在呈指数式增长，当提到怎么备份时，问题就出来了。

你肯定会问，在这种情况下如何备份？此时，你就要决定，什么东西需要先备份，什么东西可以先放在“冷库”里。“冷库”的意思是一些成本比较低的服务器，但是放在“冷库”中的数据不能随时使用，需要调出来才可以使用。

针对这种情况，有人说，我们仅把3年前的数据都放进去吧，够吗？答案就是：还是太多了。有人说，那我们可以把一年半以前的数据都放进去吧？我说不行，因为用数据观察业务发展趋势的分析师一般都要看3年的数据，所以这种做法也不现实。

在面对“决定放什么数据进‘冷库’”和“决定什么数据在紧急情况下一定要保护”的问题时，你就会发现以前我们所讲的观点——数据先收集起来，将来再使用，完全是一个伪命题。

以前，从来没有人对这个伪命题表示过异议，无论银行，还是金融机构，甚至以前的互联网公司，而当大数据出来后，这个观点就成为了一个借口、一个伪命题。

是的，这是一个很难下的决定，但这就是你必须要做的决定。如果，你在以后的某天发现你需要的数据，没有得到提前保存的话，那就只能错失这一发展机会了。事实上，这是企业的博弈。

你可能会问，一家企业并不需要从事所有的商业，为什么所有的数据都要收集呢？事实就是这样，这是数据人在管理上的不负责任，平心而论，这个责任也非常难承担。

大公司们正在数据管理这条路上学习，而当前我们面临着很多以往不曾预见的问题。比如，我们是应该在各个部门里运作，还是集中管理数据？我们是应该在数据安全的前提下更开放，让更多人找到数据的价值，还是应该更封闭，让泄露数据的可能性更小？另外，个人隐私怎么去保护？我们怎么才能成为一家负责任的数据管理公司？这些都是代表性的难题。

现在，大型的互联网公司通常都同时拥有成百上千种在开发的项目，它们都在直接或间接地改变着数据，而在这种情况下，又如何保障数据安全？事实上，数据的源头已经“脏”了，而下游使用数据的人还不知道，同时，源头的数据使用者也没有责任告诉下游这些数据已经“脏”了。

所以，我的观点是如果你数据使用得不好，这对你本来的发展影响也不会很大。但是如果你数据使用得好，而且将它作为公司的核心竞争力，那么你的麻烦就大了。因为你的数据源本来就来自各个地方，而每一个来源没有责任要告诉你，今天从它那儿来的数据是正常的和可靠的。特别是大数据出现后，数据的精准与否更加重要。因为，大数据在很多情况下，是利用外部数据来帮助内部数据进行调整的，如果你的内部数据都难以保证“干净”的话，那么外部数据同样无法保证“干净”。

数据管理，是大数据行业的“脏活”、“苦活”和“累活”，是最悲催和最难解决的事情。 如果没有这些背景作铺垫，我对很多公司在做所谓的大数据的运营就持有怀疑态度了。所以，我觉得很多所谓的“大数据专家”都是作家，而非真正的数据人。

用，从收集数据到管理数据

在“用”数据的问题上，数据的分裂和重组，都能做到颠覆性创新。

到底应该如何应用数据，才能用来解决当今公司中存在的一些问题？在“存、管、用”的过程中，如果使用时产生问题，我们会询问是决策的问题还是执行的问题；如果是决策错了，我会责问相关人员是否明确了收集回来的数据的作用，以及有没有保管好这些数据，如果不知道，那就是一个偏差；如果不是数据结合过程中出错，那就是数据源本身出了问题，就要对数据的源头“存数据”更加严格。这时候你会发现，“存、管、用”也是一个反复验证的过程。

在“存”数据上，我看到的问题是，很多人往往只看到数据的表象，比如，根据有过购买行为的用户来做数据收集，而大部分在“冰山”之下的数据就放任不管了。

我依然用某电商母婴频道的例子来解读，收集数据的维度是多种多样的。再假定如果什么“数据”都可以获取，如果中国每家医院的数据

库都是开放和可合作的话，那么这个数据就又庞大又准确。

在“用”数据的问题上，数据的分裂和重组，都能做到颠覆性创新。用户的生理性别是决策的重要依据，而现在可以从购物属性上变成“早男晚女”，这种改变并不违背运营数据的原则，因为这是购物属性，这个创意来自“本来不可以分裂的东西分裂了，本不应该重组的东西产生了新的力量”的理念。比如，你只知道用户的地址，然后根据安居客知道他房子的租金，再加上一些外部信息的补充，如10公里内有没有银行和电影院，有没有大型购物中心，然后就可以在数据世界中重现他的财富圈和生活圈。而当这个地址是办公室时，那就可以得出他办公生活区和日常生活这两个生活圈。

从使用数据的角度来说，电商行业就有很多值得其他行业借鉴的地方，可以让数据能够真正地使用起来，并且产生实际的商业价值。先来看看现在电商的背景，不论是以阿里为代表的平台型电商，还是以京东为代表的自营型电商，或者以一号店为代表的垂直类电商，他们的一个共同特点就是商品非常丰富，商品数量动辄就是百万千万级，而平台类型电商的商品数量可能更多。

对于消费者来说，进入一个电商网站的首页并不需要看到那么多的商品，如果消费者有明确的购物诉求，那么，可能会直接进入电商网站的搜索引擎开始寻找商品；如果没有明确的诉求则可能是在电商网站提供的类目和活动等区域随意寻找。这个时候问题就来了，页面内容是有限的，消费者的时间是有限的，消费者的需求是有偏好的，但是商品量非常大，电商的目标又是为了能够通过闲逛让消费者产生成交额，那么，如何找到合适的商品放在首页就成了问题的关键。

面对这样的问题，我给出的解决方案是通过一套数据中间层，来生成用户在特定市场的个性化标签。电商企业不同类目运营的员工通过算法或者人工选品来实现用户标签和商品的匹配，从而实现用户“逛”的效率最优，进而提高用户的购买转化率。

第一步，建立用户的标签

建立标签，简单地说就是通过数据的分析来对用户的偏好进行描述，建立标签通常有以下三种方法。

1.通过业务规则结合数据分析来建立标签。这一类型的标签和业务人员的经验紧密结合，这里可以举几个例子，以对这类标签的设置有更加直观的感觉。

比如，业务人员可以判断出购买某一个具体车型的人可能就拥有这款车，此时，就可以通过数据进行分类，把用户分为大众高尔夫车主和福特福克斯车主等，这个时候当用户进入汽车配件类目时，就可以直接为用户推荐相应的汽车配件，直到用户有明确的行为去搜索别的汽车用品时，再进行数据调整。再比如，有些用户平时很少网购，但一到大型节日前就会大量购买商品，这一类用户通常都是企业的采购人员，这时候就可以在礼品等类目进行企业礼品的相关推荐，甚至直接推荐该网站的储值卡。还有，对于中老年人的识别，可以通过用户经常使用的地址和包裹的寄送地址来进行区别，比如，经常在某地给一个固定地方寄送包裹的，那么结合购买的商品就可以认为包裹的接收人可能是老人，就可以给用户贴上一个“家有老人”的标签，在相应衣服类目中就会出现中老年服饰。

2.通过模型来建立标签。比如在婚庆类目上的特定行为，当然，特定行为是通过数据模型识别出来的，此时我们就可以认为其是一个即将结婚的用户，这样可以结合时间来给用户打上婚庆标签，也可以持续观察这一类用户，在未来可能会打上家装的标签和母婴的标签等。结合用户的手机充值和收货地址等行为，可以用模型计算出该用户是否是自己购买，还是作为一个网购的中心者为他人购买，如果能判断经常为他人购买，则可以打上类似于“网购影响力中心”这样的标签，可以在不同类目的场景中运用。

3.通过模型的组合来生成新的标签。任何一个模型都是有生命周期的，或者说企业内部不同的建模人员可能对同一用户会做出不同的判断，所以，我们需要对模型不断地进行整合。通常情况下，可以采用模型投票的方法来从多个模型中抽象出合适的标签。比如，在三个模型中，两个模型认为宝宝是3~6个月，一个认为宝宝是12个月以上，那通过模型的整合，应该可以确定宝宝为3~6个月。

第二步，标签的应用

标签的应用是指在电商网站的首页或者是具体的类目频道页，进

行标签的使用。

标签的使用，最核心的就是数据中间层和前台业务层的对接，并且能够让运营人员非常方便地进行商品的设置。这里涉及两个核心点：一是中间层和业务层的对接，二是中间层的易用性。下面分别就这两个内容来做一些探讨。

1. 中间层和业务层的对接。目前，对接是在互联网广告中非常热的概念，典型的应用之一就是DMP（Data Management Platform）。在这个系统中，用户以标签化的形式存在，也就是之前给用户打好的标签有了一个管理的平台，终端使用者可以在这个系统中进行用户选择，选择完成之后就会产生一个投放计划。DMP还会和前台业务平台进行打通，简单地说就是用户登录首页之后，系统就会认出用户身上的标签，就可以根据DMP中设置的计划来产出不一样的内容。

2. 中间层的易用性。对于终端用户来说，选择标签需要足够简单，并且能够非常清楚地知道这个标签具体代表的含义是什么。对于数据从业者来说，让数据变得超级简单是一个非常重要的使命，所以界面的设计和后台的管理等内容都非常重要，否则可能会失去标签系统的价值。

对于大数据来说，“用”是让数据发挥价值的最大一步，在这里我也只是举了一个数据应用的简单例子——标签系统。这个例子是数据和运营数据紧密结合的一个案例，也是数据运营或者数据驱动的一个典型案例。只有先结合大数据的技术将数据化运营做好，才能让数据从成本转化成利润，才能真正发挥出大数据的价值。

用化骨绵掌解决本质问题

绝大多数人都可能习惯了在现象层面思考问题，因为这时不需要花太多的力气思考。事实上，我们的逻辑思考是一个慵懒的系统，那么，如何找到本质才是解决问题的关键。

一方面，要找到本质问题，就需要我们在思考时排除过多的东西，使我们能够更好地集中精力去解决重要的问题；另一方面，我们要有方法去发现本质问题。

能够透过现象看本质是一个非常重要的方法，对这个方法最简单的描述就是，不断地用逻辑方法将问题进行分解，直到不能分解为止，然后从根本处去解决这个问题。

我们还是以电商行业举例，以说明如何从现象出发找到本质。绝大部分的电商，其实传统行业也是一样，关注的都是成交额——利润来源于此。但如果仅仅问一个问题——如何来提高成交额，就非常难以回答了。所以，必须把这个问题分解了来回答。

通常情况下，在电商行业，我们会把成交额拆分成几个指标——日均UV、浏览转化率（UV到达商品详情页面的转化率）、购买转化率（到达商品详情页面的用户到购买商品的页面）、笔单价（平均每一笔的金额）、人均笔数这5个指标。不过，值得注意的是，这些指标都是平均值的指标。从诊断大盘健康程度的角度来说，这些指标是可以使用的，但是如果要用来分析更为细节的业务或者应用于别的公司，就要注意数据的陷阱了，盲目套用很有可能会造成巨大的损失。

在指标分解完成后，我们就可以用这些指标来分析公司的业务了，我们把这些数字进行了图形化（见图10—1）。

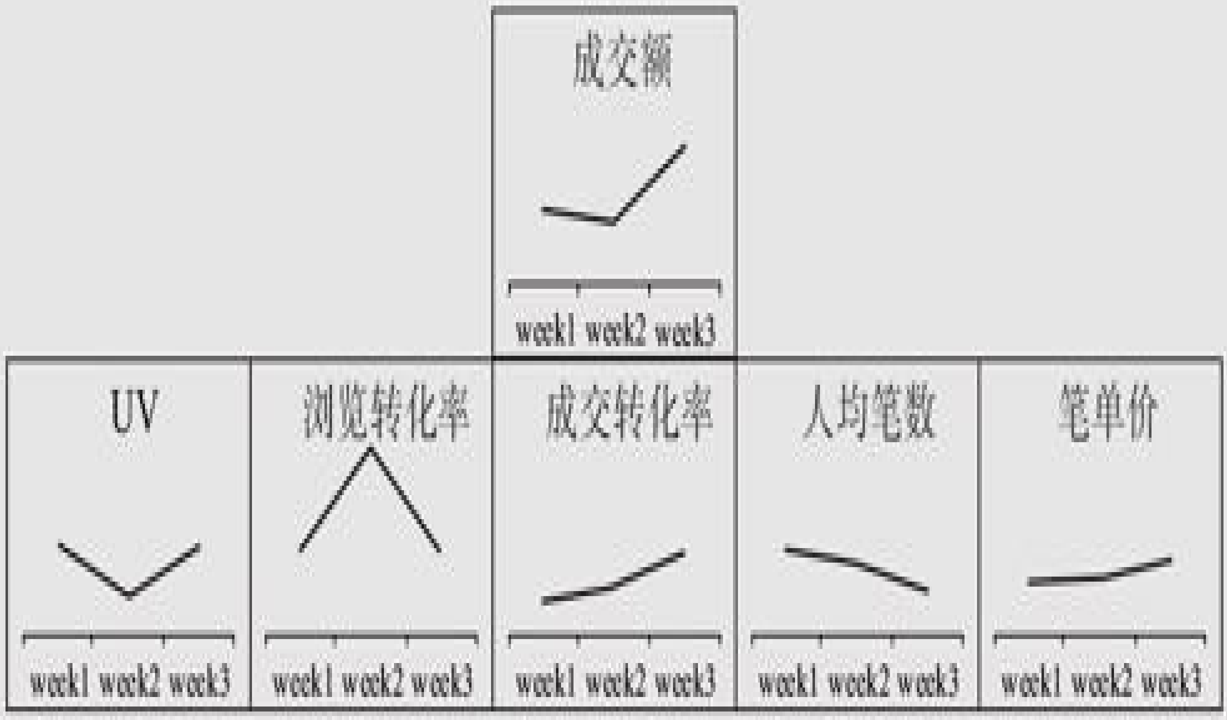


图 10—1　5 大分解指标对电商 A 第 1~3 周业务的影响

这是不是更直观地指出了问题所在？最近一周成交额上涨的原因是在于笔单价的上涨。我们再来看看图10—2，这会让问题更直观，能够使我们更清晰地看到问题所在。

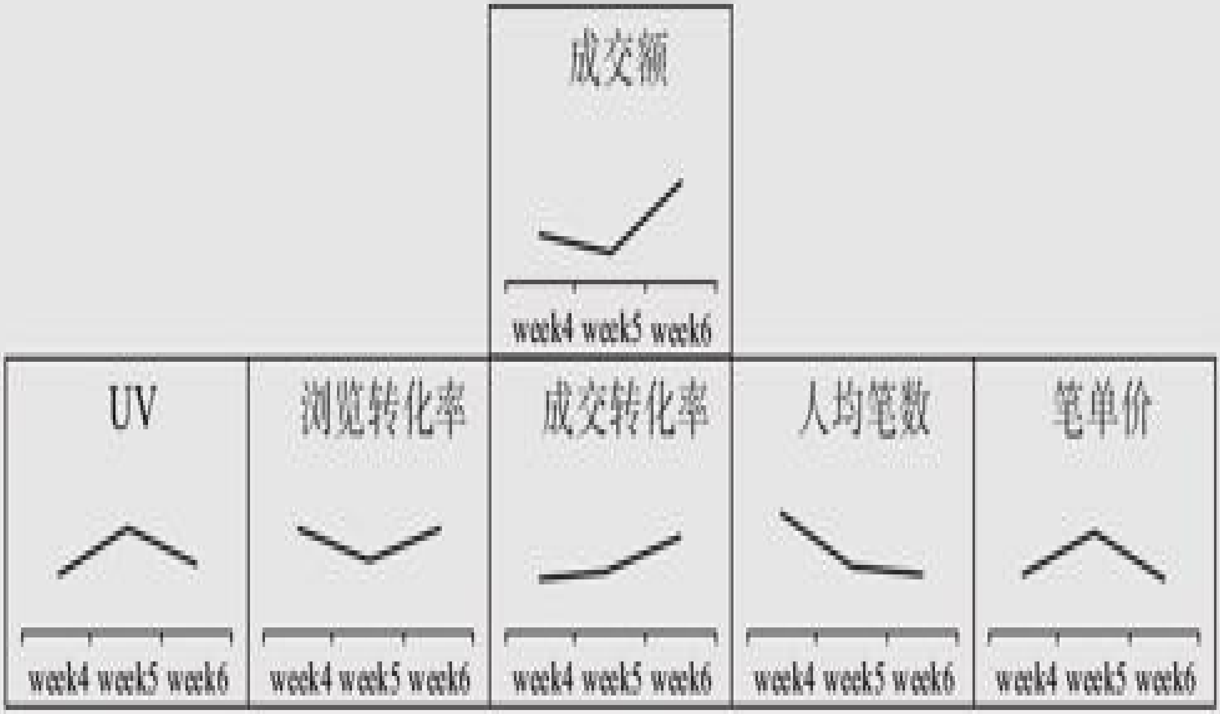


图 10—2　5 大分解指标对电商 A 第 4-6 周业务的影响

图10—2从总体交易额的角度看起来趋势差不多，但是当我们更深入地去分析原因时，发现前两个指标出现了截然不同的走势。如果我们不进行深入的分析，只是不断地去猜测到底是什么因素导致自己的销售额下降，那这可能也就只能做到猜测的层面，对企业的发展毫无实际价值。

在分析问题本质的时候，有一个很好的方法可以用，叫作MECE（Mutually Exclusive Collectively Exhaustive），意思是彼此独立、完全穷尽。这是一个麦肯锡方法，这个方法很像图10—1至图10—2这两个对销售额分析的分解图，只有在解决问题的时候要做到彼此独立且完全穷尽，这样才能够找到问题最深层次的本质原因，才能够就这个原因找到合适的解决方案。以上关于销售额的两幅图其实在不同的业务场景下还可以进一步地细分，比如UV还能够分成广告和非广告或者首次和非首次等。

总的来说，当我们遇到一个问题之后，一定要问自己这到底是现象还是本质，如果是现象那就一定要找到本质。被表面现象迷惑是很多人都会犯的错误，而所谓的成功，就是少犯错误，就是早一点从本质出发去思考问题。

11 大数据，未来商业的利器

今天，我们正处于决策成本产生巨变的爆发点，过去那些想尽办法都无法获取的数据，在今天唾手可得，而当有些表面上完全不相关的行业数据关联起来时，居然产生了新的商业价值。更重要的是，过去，我们更多地是带着问题去寻找能够验证自己观点的数据，而今天我们却可以使用数据去预测可能出现的问题。海量数据可以使人的智慧得到更大的发挥，并变得更加规模化。大数据的本质是人，数据研究的极点就是莫测的人性。我们一旦掌控了数据之后的数据，就会拥有制胜未来商业的无敌利器。

假定数据是脏的

> 在处理数据的时候，会像污水处理厂一样，每一步都问自己要如何处理这些污水。这种情况的出现，到底是因为数据源脏了，还是因为数据提炼过程做得不好？

美国有一家初创公司，专注于与地理位置相关的数据收集、整理和查询服务。它将地理位置的相关指标，按照酒店和旅馆等属性划分为不同细类，对外提供基于位置信息的实时查询，为包括美国最大点评网Yelp在内的多个知名应用提供底层数据服务。

这家公司最令人印象深刻的是，它对于所收集来的数据会提供一个数据质量评分，以反映数据的可信度和质量水平。它会对这些数据的源头以及对处理数据阶段所用的算法进行评分。也就是说，这家公司在提炼数据的每一个阶段都进行了数据化管理。

这家公司的做法让我们看到了一个趋势，也是一个非常重要的趋势。因为它首先已经接受了数据源肯定是脏的和数据源一定会被污染的事实。所以，它在处理数据的时候，会像污水处理厂一样，每一步都问自己要如何处理这些污水。这种情况的出现，到底是因为数据源脏了，还是因为数据提炼过程做得不好？这个过程我们一定要区分，而且这样的区分是可取的。这家公司是假定数据是“脏”的来做数据管理，而不是假定数据是稳定的。而且，假定数据是“脏”的来处理数据，在大数据时代将是一个非常重要的趋势。

事实上，我们今天在处理的大数据，依然只是冰山一角，而更大的数据都隐藏在我们的语言中，比如我们说的话和写的字。所以，将来我们要准确地从互动中抓取数据，也一定要依赖对自然语言的处理。现在，美国的很多数据研究人员都在瞄准非结构性数据，即语言处理这一领域。

学会慢慢淡化数据

数据是有优先值的，在数据中有些是特别核心的，有些即使缺失了也没有多大问题。所以，我们要学会真正坐下来盘点那些对公司最有价值、对用户最有价值的数据。

想要确定数据的优先值，就要先解决以下几个问题。一是数据的标准化。在大数据时代，我们需要一个标准化的东西供我们进行交流。二是我们到底如何对接和交换数据，如何在交换的时候保持数据的稳定性。比如自然语言，比如在无线和PC不同场景下受到的影响，这些情况都会滋生出许多新问题。

第三个重要的问题是数据的存储，这将涉及数据的时效性这一问题。有人曾经提出过一个很有价值的观点，即现实中，网站最大的场景变化就是网站改版。因为重新设计网站，本身就影响数据，比如公司的详情页和首页，任何改变都在影响数据。如果在1~3年后，你才说得出数据的这一改变是由于促销、用户行为或是改版引起的，那这一数据就已经没有任何价值了，这就是数据的时效性。

所以，美国出现了一个概念叫数据淡化（Data Decay），意思很明显，数据会慢慢淡化。我们要更清楚地认识到，数据是有优先值的，在数据中有些是特别核心的，有些即使缺失了也没有多大问题。所以我们要学会真正坐下来盘点那些对公司最有价值、对用户最有价值的数据，这是一个非常重要的趋势。

数据的标签化管理

数据的属性标签是人类经验判断的数据，是数据后的数据。

数据的属性标签是人类经验判断的数据，是数据后的数据。例如，当你要为一件物品打上标签时，其实就已经动用了你的经验数据分析，并进行了归纳总结，结合当下的环境给出了判断。如果没有考虑环

境的影响及准确性的评估，这种经验加上直觉的判断是不稳定且又难以解释的。但从数据收集的角度去看，数据的属性标签又是一个潜力极大的数据。

在数据属性的管理上，对于用户来说，每个人身上贴的标签都是多种多样的，但是对于企业来说，如何将这些标签归一，如何用一个点去将之串联，又如何把这些点连起来去描述这个用户，这才是核心问题。

比如说，你要去应聘一家公司，A公司对你的评价是“很可靠”，B公司对你的评价是“不可靠”，而C公司准备雇用你，然后C公司看到了A公司和B公司给你贴的两个标签：“可靠”和“不可靠”，它就会困惑。

这样的问题常常发生，那么，我们要怎么去做标签的管理呢？

首先，我们要明确的是，“可靠”这个概念是没有标准化意义的，除非在定义标签之前，我们就界定清楚这个“可靠”的标准是什么。否则，“可靠”的标准是准时还钱、说话算数，还是他向来都很守时，我们就无从得知了。如果这些标准是**“**可靠**”**，就给予了我们一种可以还原数据的能力。所以，在属性管理中，假如属性是“×”，那么我们一定要定义清楚什么是“×”，在没有清楚定义的情况下，这个数据的属性是毫无价值的，而且，将来你也依然不知道怎么使用这一数据。

标签在观察之后加进平台和直接加进去是不一样的。在电商平台中，就有一些标签是在观察后加进去的，如果由卖家自由地加标签进去，必然会造成混乱。所以，标签的属性管理，在运营数据中非常重要。

标签化管理，是一个非常重要的趋势。电商企业今天面对的一些问题在美国的电商企业中同样存在，可见，我们发现的问题，别人也在面对，不过这些问题并不需要现在就去解决。

属性管理的层级化十分有必要，但是在使用数据前，必须要了解数据的场景、数据是如何放进去的和数据的场景是什么。在这一切未知之前，就说数据如何好用的话，是不可能的。所以，现在企业运营数据的趋势是，我们应该找出一些属性进行归类，然后再慢慢地考虑如何提炼，这对于未来非常重要。

重要的是数据和数据之间的关系，而不是数据本身

大数据价值的实现，在于数据与数据的连接。

Google做了一件非常惊人的事情——Google甚至能在不明白某个网页语言的情况下，知道其内容是什么。试想一下，如果你懂俄语，看出俄语网页里在讲什么当然很简单。但是，如果你仅仅通过看字词的排列和网站的分类，就知道网页的内容，这是不是很令人惊叹？

这就是知识图谱，它是一个无穷无尽的世界。事实上，知识图谱并不是数据，而是数据和数据之间的关系。但这里有一个非常大的弊端，就是数据的储藏量非常大、储存的方法也很复杂，且稍微改变一点点关系的定义，整体就会产生巨大的变化。

比如说，有一个知识图谱在说电商平台用户之间的关系，那数据信息就非常庞大了。试想一下，今天电商平台里有多少个用户跟你有关系？假如说有25个人，那么25个人的关系就演变成了25×25条关系。这时候，我再问你“什么是关系”、“见过就算关系，还是一起买过东西叫关系”的问题就具备了一定的难度。

关系建立的维度是无限大的，而且定义稍微改变一下，整个存储和整个数据库都会发生变化。所以，知识图谱的把控是有难度的。举个贴近我们生活的例子，比如说银行很早之前就给你开办了信用卡，决定银行这一决策的不是你的个人关系而是总关系。银行决定是否贷款给你，是要看你爱人做什么职业以及你家中其他人的经济情况如何。当这种种关系关联起来时，就会产生一个极为重要的知识图谱。

以往我们谈大数据时候的本钱，莫过于“我有这种数据，你没有”。在未来，数据和数据之间的关系才是重中之重，而不是单纯的数据本身。

数据的实时化与实时性分层

我们千万不要把所有的能力都用来处理实时化的问题，因为我们依然会有大量的数据需要在恰当的时机（**Right Time**）处理，有的数据是重要的，但不紧急。

以上讲到的很多内容都是关于数据收集和管理层面的，而在数据

的处理上，我在LinkedIn上看到了一个很有趣且有价值的做法。LinkedIn在处理数据时，会把公司的数据服务分为几层，一方面是紧迫度，另一方面是重要不重要。比如，它会把数据分为“快数据紧急”、“快数据不紧急”和“慢数据重要”等。

我觉得对数据实时性分层的做法是合理的，而有人觉得数据的实时化处理是趋势，但是我持有一定的怀疑态度。Real Time是“实时”，Right Time是“恰当的时机”。但是，据我看来，数据处理不一定要实时。比如，我们常见的情况是，每家公司都有财务的相关数据，这类数据的处理都是“T+1”，意思是你想要的数据在第二天才可以拿到。因为在其他数据没有到位的情况下，数据实时化的价值也不大。

但是，换个场景来看，银行若要判断某个人的信用卡是否被盗用，那么肯定要对之进行实时化处理。数据的实时化，让我们从商业的角度去认知数据，值得注意的是，在具备了实时化的数据处理能力后，很多以前不能解决的场景开始变得能够解决。在未来的某一天，编码的工程师能够在编码时直接写上“如果一个三天前只浏览未购买的客户回来了，我要不要给他一个两元钱的红包”。这个程序是完全编好的，用户登录本身就成了一个实时标签，快速的运算会让每个网站都具备最强的时效性。

我们再换一个角度来思考，如今手机、电视、游戏机和PC等多屏运作的时代下，作为一个网站，有多大的能力在非常快速的情况下，让自身在非常小的时间点里抓住消费者，卖出产品，这种实时的能力会在未来的商业中变得越来越重要。

一个网站必须要让自己的实时能力更高，甚至用户接下来的每一步你都应该猜到，但我们千万不要把所有的能力都用在处理实时化上，因为依然会有大量的数据需要在恰当的时机（Right Time）处理，有的数据是重要的，但不紧急。

未来是人机的结合体

人和机器的结合，或者人和数据的结合将是未来的一种进步模式，人类将通过数据变得更加智能。

很多人会问，大数据目前发展到什么阶段了？我的答案是——水

分太多的阶段。但毋庸置疑，大数据已经极大地影响了我们的社会，但还远远没有达到它的爆发点。因为有了大数据，人的经验开始和数据结合，两者相互激活，让人的智慧得到了规模化地放大，这也使得整个社会开始伴随着数据的发展产生巨大的改变。但是，在大数据的使用上，未来的发展空间注定会比现在取得的成绩更加宏大，“数据化运营”和“运营数据”这个数据与人之间的闭环系统也会越来越完善，人机结合仍然有巨大的空间让我们去挖掘。

未来，数据的种类将超出我们的想象。以前的数据更多集中在外部行为的监控上，不论是网络购买行为，还是网上社交行为，都是用户在“远距离”提供数据，即便这样，我们仍然还没有运用好这些数据。随着可穿戴式设备的出现，数据和人将真正融为一体，类似谷歌眼镜这样的设备，将让我们看到的东西即时数据化；类似健康手环类设备和可以深度收集脑电波数据的设备，将随时会使我们人体的活动转化成数据。目前，记录睡眠状况只是其初步的应用，在不远的未来，用数据记录我们每一秒钟的生活也将成为可能。

当万物皆数据的时刻到来时，商业发展的更多新机会将会出现，数据将会帮助我们更好地做出判断，比如，什么时候最适合吃饭、什么时候身体疲惫适合睡觉和什么时候记忆力最好等，这些都能通过数据来进行预判。甚至于，当记录了人的足够数据之后，数据就可能告诉我们此时此刻应该做些什么及最佳的策略。也许在那时，决定人是否聪明的指标，已经不是IQ，而是是否拥有足够优质的数据。

人和机器的结合，或者人和数据的结合将是未来的一种进步模式，人类将通过数据变得更加智能。

最后我想用两句话来总结：

当下，我们要学会人机分工，让人做人最擅长的事情，让机器做机器最擅长的事情；

未来，我们要相信人机结合，人机的界线已经模糊，无人驾驶汽车已经变成可能。未来人类的身上流动着的是血液以及数据。

忽略了趋势，过去的价值一文不值

我们通常都认为，经验应该是越多越好。一个有经验的人总是更能够在很多事情上做出正确的判断，因为积累使然。但事实上，我们发现，有时候经验越多，似乎越容易让人犯一些低级错误。这就好像，你会发觉往往天天研究彩票和股票的人总是发不了大财，而往往是一个菜鸟莫名其妙地就中了大奖和买了一只牛股。

为什么经验有时会使人犯低级错误呢？在这里，我们应该将其分成两种情况来看待：一种是信息不对称；另一种则是逻辑错误。

当我们讨论信息的不对称时，首先要讲的就是经验。经验是对过去的度量，但不是所有经验信息的质量都很好。在经验的数据库里，肯定有一些信息是正确的，有一些是错误的。当经验中混有很多噪音干扰时，我们会跟随错误的经验做出判断，此时，我们就会发现自己变笨了。

比如，在使用没有评估过的经验时，你用A方案获得了成功，用B方案却失败了，而且在评估的过程中也不是单纯的0或1，而是用0~1的范围来度量的。说到这个，就要提到数据分析师了。一般而言，数据分析师对于信息是很有洁癖的，也就是在对于经验的累积上，他们对质量把控得十分严格，如对于数据信息的排序、分析可靠的信息源进行多次使用、了解信息的出处和知道信息的提供者等。根据这些，数据分析师在它们的辅助下做出了决定。这意味着，你所有的信息来源都需要有正确的途径和渠道，不然，这个决策也会出现偏差。

除此之外，还有一个因素导致经验使人变笨，那就是在分析时忽略了趋势，这个道理可以用赛马的故事来举证。

在香港，你经常会发现拿着一大堆材料的人会分析马匹的数据。比如，有人会根据一匹马进行1 200米跑的时间，来计算它未来可能跑完全程所需要的时间。但我们发现，还是有很多人因算错导致赌马失败。为什么呢？因为历史数据和我们今天面对的情况中出现了一些假象。香港大多数赌马的人，他们最终收集到的数据都是受到影响的，而不是经过清洗的，当然是不准确的。每一个赌马的人都在看过去的数据——马会会给每一个赌马者提供前三场赛马的数据，大家只会关注这个结果，而不会去关注赛马当天发生了什么。如果是我，则会去回看录像，就可能发现其他的情况。比如，如果这匹马本来想发力，但前面有马匹挡住了它，它才被扣除了两秒钟；或者骑师扬鞭，鞭子掉了，扣除5秒钟；再或者有些马发脾气偏离跑道，也要扣

除秒数。当排除所有意外算出的时间，就是干净的、没有影响因素的真正经验了。这时候，我们得到的第1~3名的数据和最终比赛结束公布出来的第1~3名的数据就会不一样。

除此之外，还要观察有关赛马成长的趋势问题。事实上，在每一场赛马里都会出现很多意外，如果这些意外都不出现，也会直接导致结果的不同。因为每匹马幼年和成年的状态是完全不同的，因此它的数据价值评估也是不一样的。一匹马3～4岁等同于我们人类18～25岁，体能上会出现非常大的变化；又或者，年幼的马匹对于骑士的体重非常敏感，多1～2磅就会有非常大的影响，但长到5岁后，负重对其的影响则会慢慢降低，这就是硬趋势。即便有很多过往经验的数据都不能很好地匹配今天你要做的事情时，意味着经验需要将数据进行清洗——把当天的影响因素都找出来，并还原，这时候得到的数据才是正确的经验。

结论就是，经验使人变笨的原因在于你之前的经验本身就存在误差，即数据源本身就存在问题，而这种误差一般人看不出来。因此，在过去的经验积累本身就有问题的情况下，根据经验得出的结论自然会使人犯低级错误。

你永远不要假定这个世界是真空的，所以你还需要多多观察频繁出现的新数据。当有新的数据出现时，以往的经验就需要重新做评估了。

结 语 开启属于你的个人大数据管理

BIG DATA 众说《决战大数据》

The Business Revolution

徐 雷 京东集团高级副总裁

老车是“手艺人”，数据时代解读数据世界奥秘的手艺人；对我来说，亦师亦友、亦火亦水。

我们曾若干次品茗促膝、围炉夜话，人生的起起伏伏、数据的理解与应用、佛法修行，每个话题始于数据终于数据，获益匪浅。难能可贵的是，我们对数据的应用、理解，观点出奇的一致，平添相互好感。

老车是数据的修行者，我十分期待这位无疆行者的感悟与升华之作，一杯清茶、一本好书、一位知己好友，人生幸事。

过去几年，数据分析师们问我最多的问题是，如何才能具有全局观的思考能力，洞悉一切。而普通人问我最多的问题则是，我如何才能保持幸福快乐的生理和心理状态。

这一切的答案要从“命运”开始说起。

我一直认为，“命”由天注定，而“运”是可以改变的。人生绝非宿命，公司和个人并无二致。

人一生下来，身体素质、DNA和天赋往往都是注定的。而“运”呢？却是不断的选择和决策所组成的。

“命”就像一辆离开流水线运出工厂的汽车，它的质量和耐用程度从走下流水线时就被注定了。“运”却是这辆车后天所开辟的道路，大路与小路不同，每一条路选择得对或者错，对人生的影响或大或小。有时，小路也能使人生出现极大的转折。机会错过了永不复返，挫折也未必不能避免。

在我看来，成功路上，有**4**种东西是可以积累的，分别是财富、人际关系、知识以及思维方式。前两者，大多数人都能够直接明白。比如，没有钱很多事情都办不好，多个朋友多条路。第三点知识说的是人从小学、中学、大学以及进入职场获得的各种知识和经验。在现在的职场上，每个人的竞争能力及赚钱能力，大都和知识有着直接关联。

而不被绝大多数人知觉的是第4点——思维方式。思维方式与你所做的选择和决策密切相关，冥冥之中决定了“运”的走向。当下，我们生活在大数据时代，各种数据充斥在我们的生活环境里，每天都能收集到来自自己所见的文字、影像、声音等，以及来自各方面的信息。而在这样一个信息泛滥的环境里，很多人完全没有关注过如何积累、分类及使用当前越来越容易获得的大数据，更别说让大数据变得能为个人所用。

数据的积累，从收集到还原

有的人每天都会在同一个十字路口迷路，也就导致其每天都会迟到10分钟。那么，为什么会出现这样的错误？这是因为他根本没有重视数据的积累，也没有想过使用这些数据去产生什么价值。在我们之中，绝大多数的人对事物的态度也只是观察，却不会思考。而我的做法是，我要对数据做一个积累，我要对我所在行业中呈现出的信息进行积累。

这是一个有限时间和无限大数据进行博弈的时代。如果没有积累数据的意识，没有培养自己数据化思考的意识，改变命运真的就成为一种投机行为。

古语有云：“人无志不立。”如果你有志向，你就会知道自己应该积累什么样的知识、什么样的财富和什么样的关系。如果你进行了大数据管理，到了某个阶段你又有了很好的数据化思考方式，你会发现财富、关系和知识的积累完全能够为你所用，到那时你的人生就会有很大的不同。

至此，你应该已经感觉到了这本书前文所讲的大数据内容，其实完全是可以为个人所用的。我可以用我的经历来举例，我相信，如果你现在开始用大数据的方式来进行个人的知识管理，开启个人大数据的积

累，不超过6个月，你就会发现自己开始变得大不一样了。

做好个人大数据管理

我现在每天早晨四点半起床，熟悉我的人都知道我每天早上必做的一件事情是“品觉晨读”。做“品觉晨读”这件事情大概有4年了，它是我个人大数据管理中的数据收集“工具”，就好比运营数据时的“养数据”概念。

今天，如果你开始关注某一领域，想对某一种知识进行积累，在开始养数据的过程里，你会如何开始呢？这就是一个棘手的问题。没有人能够在第一次做运营数据的时候，就知道这个数据是好的，是对自己有用的，或者，什么数据是不好的，是对自己没用的。就像开始做企业大数据时，我根本不知道哪些数据一定有利，或者在未来可能对我产生重要的价值。

所以，我一开始会把每天关注的那些数据收集回来，放到我的Evernote里面，为什么我要选择这一笔记软件呢？因为它可以方便我做数据管理。

- 它能够方便我用各类终端，包括手机去收集数据；
- 它能够把每一份收集回来的数据进行分类及打标签(结构化)；
- 它能够方便我将数据随时随地地调用出来。

时至今日，在我积累的数据中，已经有超过上万篇关于过去我经验中的PPT及早上收集的电商新闻，如果我没有一个非常好的分类/标签管理体系，我如何才能做到这一点？所以，要完成这个大数据积累的闭环，要求放进去的数据，一定要能够非常快速、实时地调用出来。我一直在训练自己怎么使用两三个关键词就能够让我在10秒钟内调出我积累的有效数据。当我想要使用数据帮我做决策时，有效的数据就在手边，这个时候运营数据，即养数据，就已经产生了作用。

现在，我每天依然在继续收集数据，在这儿我可以透露一下我自己的一个数据收集的“狠招”。在我的微博关注里，我按照两个维度来分类我关注的人：一是他们懂什么东西，即有什么特长；二是根据重要性，进行5～20分钟，或者45分钟关注度的划

分。这样划分的意思是，如果今天早上我只有5分钟，那我就只看重要朋友发布的信息；如果我有10多分钟，我会把重要性为5分钟和20分钟的朋友发布的信息都看完。

从收集数据的角度来看：首先，我在选择有效的数据源；其次，我有重要性的选择，所以我能很好地控制我的时间。每天我做完这些数据收集的时间大约是一个小时，时间久了，我现在对每个数据源甚至提供数据的记者的能力都了若指掌。某一个学者的观点是怎么样的；腾讯的记者能力如何；哪个网站周六日是不用关注的；凤凰网比较勤奋，周六是有人上班的；腾讯网最努力，每天都有新闻。甚至，各家媒体参加了哪些会议，这些我都知道。

我现在已经对数据积累与决策质量之间的关系具备了一种直觉，所以我开始对电子商务有“觉”了。这一过程就如同前面所说的内三板斧中的 “混、通、晒”，我每天“混”在电子商务的环境中，不断地吸收和寻找规律，这是“混”；每天我将收集来的数据和我每天所处的电子商务场景进行打通，则是“通”；至于“晒”，就是这些数据如何影响到了我在工作中的决策。

2013年年底时，我听到一些数据分析师在讨论年报，他们不断地在讲2013年阿里巴巴的年报。我问他们有没有看新闻，我们的竞争对手最近有什么新动作，做了什么？京东的CEO刘强东想做什么？新闻是真的假的？你们关注了吗？所以，他们孤立地讨论这一报告是没有意义的，因为他们并没有将场景融入其中，只讲阿里巴巴自身，报告的意义也就无从谈起。

当然，在运营数据和数据化运营的这个闭环里，我们一定要假设一切都在变化，一定要去关注还有没有一些新的数据收集源头，现有的数据源会不会出现问题，我还有什么新的数据源头是缺失的。

比如，近年来兴起的虎嗅网这样的新媒体，比如199IT，比如特别喜欢发表大数据文章的人。我作为一个所谓的“大数据专家”，怎么可以不知道这个领域里谁有料、谁没料？谁是人云亦云的呢？

这就像和行业内的高手过招，我们在积累的过程中是在动态中不断迂回地走向目标的，是不断地衡量这个行业中的变化的，就像李小龙一样在和高手博弈。

个人大数据管理的挑战

在“养数据”的过程中，关键词标签是进行数据管理最重要的工具，透过标签我们可以快速定义一篇报告的内容，而数据每次调用的顺畅情况则可以用来衡量这一标签是否有效。

在我们的个人大数据管理中，最常见的关于标签的矛盾点是，你给某篇文章贴上了标签，但这个标签却是一直动态变化的。过了一段时间，你会发现这个标签失效了，而你设置标签的时候却是正确的。举个例子，你给某个朋友贴上的标签是“美女”，但是过了10年后，她老了，你这个“标签”就失效了，就又要进行升级了。我们把这个现象叫作标签的生命周期。

你以前的标签是对的，但是经过一段时间之后，这个标签就失效了，所以你需要重新更新索引，而这个过程所花费的力气非常大。多年来，我不断地去寻找对应电子商务有关的标签，以下是我总结的3点经验，可以分享给大家。

- 越有效的标签就越能让我快速地调取数据。

  标签的用途多样，比如做知识分类的、识别来源的、情景和人物描述的以及表明时序的，等等。

- 要注意知识范畴的培养。

  例如，在电子商务中，我运用了各种分类标签， 其中有行业标签：B2B、B2C、C2C、外贸B2C、外贸B2B，最新的还有O2O；技术支持方面的标签有：大数据、营销方法、 用户体验、 微营销等。

- 场景的标签我把它分为公司与人物，再加上时间。

可见，由时间链、共识面（公司、人物、来源）、 知识体系（标签）和格式（内容）所构成的一张知识图谱是解码决策分析的依据。表12—1是我做数据分析时使用的知识图谱的部分内容，以供参考。

数据分析实例

收集时间	公　司	人　物	来　源	电子商务	格　式
过去两周	淘宝	马　云	新浪科技	B2B	行业报告
3个月内	天猫	马化腾	Techweb	B2C	数据分析
6个月内	京东	刘强东	腾讯科技	C2C	新　闻
6个月外	一号店	陈　年	艾　瑞	O2O	
	唯品会		易　观	物　流	
	凡　客			移　动	
				社会化	
				大数据	

表12—1　各大电商平台的知识图谱

试想，你正需要马云一篇关于物流的文章中的数据做参考，你回忆起来这好像是上周的某条新闻。所以，要调用这篇文章的维度就是：过去两周＋马云+物流。

再增加一个维度去看的话（共识面），就可以用“过去两周＋物流”去看在这两周内还有谁提到了“物流”。

你会发觉数据标签中有些是比较历久弥新的，即生命周期比较长的，比如人的名字，不论这个数据从哪儿来、在哪里、谁说的，都不会对数据标签产生影响。但有些标签是对趋势的界定，则需要经常地总结或细分。

收集到的数据我都会如上述这样贴上标签。当要调取数据时，我会在心里问一下数据的来源是哪家媒体、哪个人、什么时候获取的。这样一来，时间线和来源就成了我锁定想要数据的重要线索。

从另一个角度来说，当我想知道亚马逊在物流中的新动向时，我会先锁定物流行业和公司标签，再配合时间线做决策。如果我想到“大数据+周涛+3个月内”，我根本不用想文章的细节，就可以将之调取出来。这就是大数据让我们已经发生的改变——我们只要记住几个标签作线索，就可以轻松地调用一万篇文章内的数据。

但千万不要觉得你可以一劳永逸了，因为这些标签会随着时间变化而“失效”，这样数据管理就成为了重要的一环。数据分析师都有变态的数据洁癖，他们对数据的来源极为看重。所以，我将人和公司进行区别对待。人是特别的，可以离开某家媒体和公司，是流动的，但是如果我将数据管理定位在人之上，就是表示我认为人更重要。如果定位在了公司和媒体的背景之上，那么公司或媒体就更重要了。

这就是变化。在大数据的世界中，没有人要求你懂得细节，没人要求你成为一名数据分析方面的专家，但是要求在你需要拿到一个专家的数据时，能够快速地调用出来。做到这样的话，我们就已经突破了人类短时记忆的短板，你就会成为一个很成功的人——真正做到了。

这也正是过去4年里我做“品觉晨读”的原因。“品觉晨读”一方面是便于我自己做个人大数据的知识管理，另一个目的就是提醒我这个数据化运营和运营数据的闭环系统不能出现中断，提醒我要坚持不懈地把这个循环维持下去。而且，“品觉晨读”也在不断地给我反馈，比如，我发完微博后会有多少人转发、多少人评论、多少同行专家帮助我一起优化我的大数据内容和标签。这使得我做数据收集的时候，也就了解到了数据收集源头的重要性，通过借助别人的力量让你的数据收集变得更加有效，即在社会上过滤。

当然，“品觉晨读”的文章只是我所收集文章的一部分，在我经历过多次数据化运营和运营数据的循环之后，我可以快速地知道某篇文章是不是对我有价值的，这加快了我收集文章的速度，但是这个快思（加快判断）也同时很容易出现盲点。为了减少盲点，每当我怀疑自己的判断时，就会列出一部分可能出现盲点的文章放在微博里，其他人自然会告诉我哪些是有用的，然后我会再观察这些数据，甚至还可能从中找出新的数据分类或标签。

人人都能成为数据分析师

个人的大数据管理，我已经做了4年，过程中困难重重，但就像大数据管理一样，不能因噎废食。我保证，只要你开始用这个方法去做，在3个月后，最晚也就是6个月之后，你会变成另外一个人。

《思考》（　　　　）中有一个例子和我的思考以及我这4年里所做的事情极为相似。书中说，如果你把某个人的左眼蒙上三天，三天之后，从医学的角度来说，他左眼相关的大脑部分，会发生退化。这个意思就是说眼睛与大脑是互相影响的，而你看到的将影响你的思考，反过来你的思考方式又影响着你看到的东西。

所以说，如果你6个月内不断地积累知识，你也会发现你的思考结构发生了改变——不仅仅是知识量上的改变，而且是思维方式的颠覆式改变。这是因为你收集数据的方法已经不一样了，你积累的东西也不一样了，你变成了另外一个人，你会觉得你“混”得特别好。

也许你已经跃跃欲试了，但是我从这几年的经历中提炼出了一些建议，希望能值得你关注。

首先，你所选定的题目，一定要是你感兴趣的，或者跟你的职业相关的。比如你喜欢宠物，你要成为养猫的专家，你就会每天收集养猫的数据，当别人问到怎么养猫时，你就能很快地知道某个专家说过相关的内容。

其次，个人大数据管理和做大数据一样，一定要从小处着眼，比如说你的目标是要成为电子商务专家，这个目标在一开始就显得过大了。从我来说，一开始，我只收集海外出口的相关资料和数据，在这个小领域里形成了一个小的大数据循环，然后走熟这个循环，再涉猎手机支付行业方面，然后是C2C和B2C行业等。在经历过小的循环之后，我才允许自己去观察别的领域。现在，在别人眼里，我是一个数据运营的专家，其实我也是这样“混”出来的。即便是现在，我也在一直训练这个循环。

再者，就是尽量让数据的收集变得简单。当你发现有用的知识后，一定要跟自己说，不要多做思考先收集，现在就收集，不是明天，更不是后天。有人在讲PPT，有价值的内容立即拍下来，而不是待会跟他索要PPT。知识的收集永远都是“现在时”，而且这个操作越方便越好，越快越好。我的收集就很简单，设置一个热键一下就能够将数据抓

取过来。

最后，你一定要把这个“运营数据—数据化运营—运营数据”的循环打通成闭环，也要利用今天社会化的优势，学会借助别人的力量。我认为这才是真真正正的大数据管理——形成一个有效的循环，形成一个可以不断获取数据和反馈的动态循环。可见，个人大数据管理能够为个人产生力量，让知识形成积累，成为你在人生道路上，从“命”到“运”最重要的分水岭（见图12—1）。

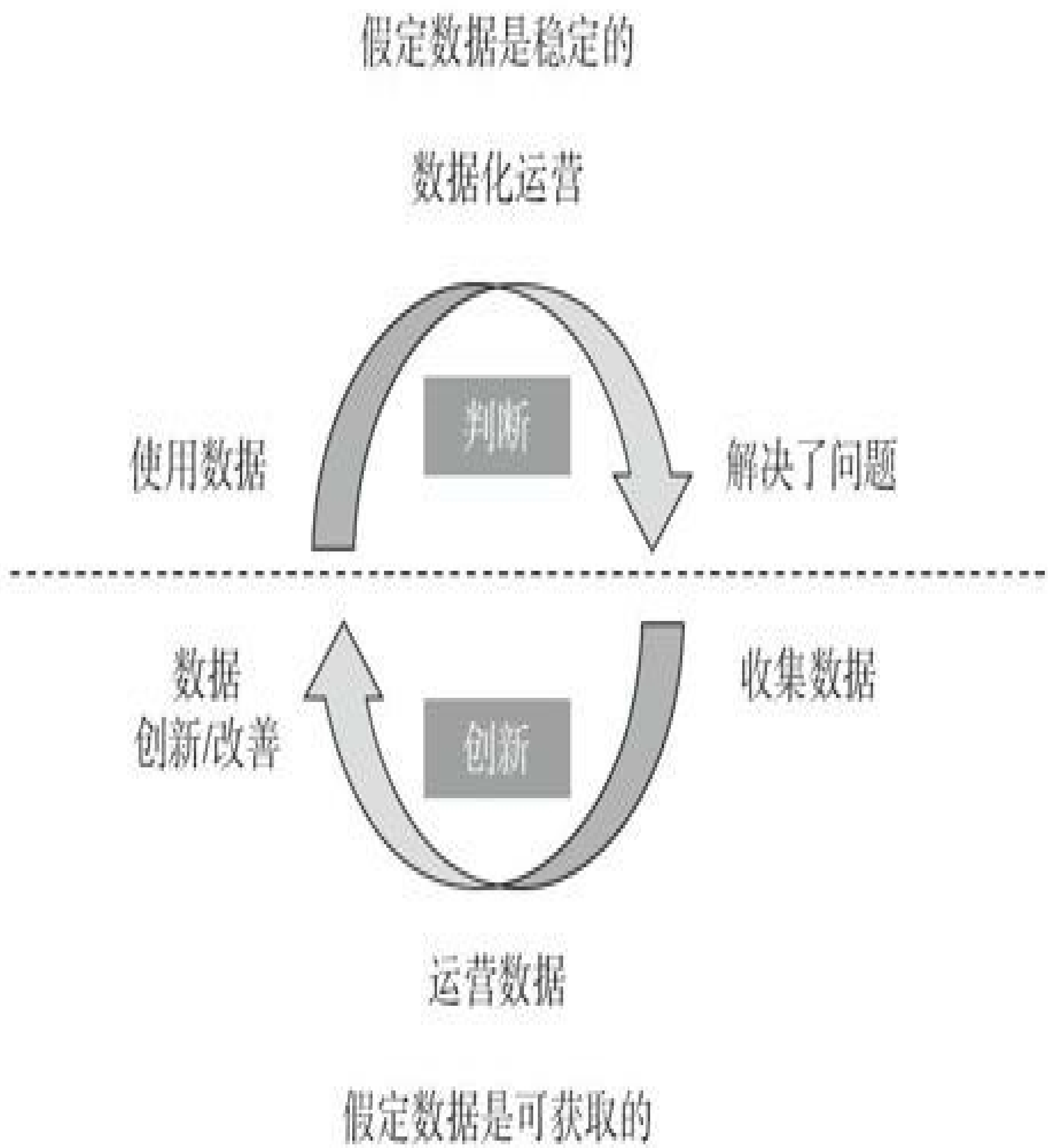


图 12—1　数据化运营和运营数据的闭环系统

现在，除了我自己这样做知识和数据管理，我还想让我的助理也加入到这个循环系统中，只要这个方式的方向正确，做好了协同，我甚

至可以找8个助理一起用同样的方式转动这一循环。而我，只需要做一个步骤：问他一个问题时，我该用什么关键词。我就相当于新建立了一个知识体系，建立了一个新的循环，我和我的助理是我们这个圈子大数据知识的协同者。

无论如何，我们的人生是要想尽一切办法来积累知识。首先要明确积累的目的，之后才是积累的方式，最后才是如何调用积累的东西，以及明确其是否影响了我的决策。只有对这一数据有感觉的时候，才能对自己的职业产生影响。

这样下来，我坚信在两年之后，你就可以成为某个领域的专家，而且这套方法论是可以被复制的。从今天开始，如果我喜欢猫的话，我要成为一个精通如何养猫的人，我只要按照这套方法论去做，我保证两年后，大多数人没法在相关的知识领域跟我相比，因为我有一套积累知识的系统方法。如果这个例子得以复制，就将会有一大群人跟我一样，那我就创建起了一个联盟。第一天，收集、决策、行动、管理；第二天，继续收集、决策、行动、管理。每天都这样实施，开启你个人的大数据管理，改变自身的知识积累和思维方式，成为一个能够掌握自己未来的人。

后 记 像李小龙的格斗一样去思考

每个男人在小的时候，心里总是会住着一个武打明星，也会或多或少地做一些武侠梦，幻想着自己能够天下无敌。在我小的时候，崇拜的偶像就是功夫巨星李小龙，他不仅是一个电影明星，更是一个武术大家，他用他的功夫开创了一片武侠的天地。

李小龙对功夫曾经有过这样的描述：

> 功夫是一种特殊的技巧，是一种精巧的艺术，而不是一种体力活动。这是一种必须使智力同技巧相配合的精妙艺术。功夫的原理不是可以学得到的，它好像科学一样，需要寻求实证，而由实证中得到结论。
>
> 在经过四年严格的功夫训练之后，我开始了解，也感觉到柔能克刚的道理，也就是如何消除对手的力道，减少自己力道浪费的手法，而这一切都必须先求得气定神闲。话听起来很简单，但实际做起来却很困难，一旦和对方交上手后，我的思想就再也难以保持清明而不受扰乱的状态。尤其是在对过几招之后，我就忘了柔的理论，唯一想到的只是不管怎么样我也得打赢他。

“以无法为有法，以无限为有限”是李小龙创立格斗术“截拳道”的纲领和要义。他认为这种方法不特别固定的格斗术的精髓就在于——在动态中制胜。他认为，稳定应该是在动态中的稳定。而之所以要保持动态中的稳定，则是因为任何的格斗都充满了不确定性，如果只能做到特定环境下的静态稳定，那么在瞬息万变的格斗环境中一定会出现逆转的机会，而这个逆转通常就是失败或成功的关键。对动态中稳定的关注，目的就是希望能够把在环境中不确定的东西变得相对确定及有规律。

如果我们把格斗泛化一下，则可以将其推广到商业的博弈中；如果把李小龙的格斗术泛化一下，则会变成一个非常重要的商业方法论。

忘掉必杀技

很多人在刚刚开始学习一个全新内容的时候，总会很崇拜老师高超的本领，常常会迫不及待地问他们：“有没有什么能让我快速达到您这样水平的绝招呢？”答案是否定的。从经济学的角度分析这件事情，你也会发现这个世界上其实不存在轻而易举就能获取的绝招。试想一

下，如果轻而易举就能获得绝招，那结果是每个人都可以获得所谓的绝招，当人人都有绝招的时候，绝招就再也不是绝招了。

在我和一些电商讨论运营数据的过程中，也经常会被问到类似的问题，比如，“我的首页该如何改版”、“我这个按钮应该放在左边还是右边”和“我该观察什么数据”等。其实，这些问题都是追求必杀技的一种问法，而我可以非常肯定，这样的问题都是没有一个固定答案的。所以，从现在开始，忘掉必杀技吧，因为世界上永远没有必杀技。

如果没有必杀技，在与高手过招时该用什么招术呢？招术就是针对对手出的招反制他的招，这就是有用的制胜技。所以，当我被问到首页该如何改版时，我肯定没有绝对的答案，我可能会问：“你们公司的首页主要起着什么作用呢？”当我被问到按钮是放在左边好还是右边好时，我会接着就问：“放在左边和放在右边的区别是什么呢？”当我被问到企业该观察什么数据时，我的问题就会变成：“你看数据的目的是什么？”只要问我的人给我的答案是不一样的，那我给出的建议肯定也是不一样的。如果我得到的答案还不足以让我获得足够的信息，那我就会针对答案不断地提出新的问题。

为什么我要不断地去问问题呢？原因很简单，就是我想知道问题到底出在哪里，以找出每一个问题背后不一样的原因。这个过程是我在思考问题时常会用到的方法论，我把它称为“迂回”。

在迂回中寻找落地点

迂回，不是直奔一个目标，而是通过不断地寻找落地点慢慢行进。可能某个目标看着很简单，似乎马上就能实现，实则可能是困难重重。在格斗中，每个人都知道击中对手要害的重要性，最终可以以“KO”获得胜利，但是如果直接出重拳击打对方的头部，那一定会被对手防死，反而会使自己露出很大的破绽，让对手有机可乘。所以，很少有比赛能够出现瞬间击倒对手的现象（除非强弱分明），高手们都是在不断的尝试过程中发现机会的。

有一次，我在和一个电商的管理者交流，谈到了购买流程的问题，他问了我“该突出‘立即购买好’，还是该突出‘加入购物车好’”的问题。他之所以在思考这样的问题，是因为他认为购买流程对电商来说是

至关重要的，而之所以有这样的困惑，则是因为他发现了两种不同的购买流程。在宝贝的详情页上，淘宝网突出的是“立即购买”，而京东和亚马逊突出的则是“加入购物车”。这就让他十分不解，因为淘宝是最大的电商，之所以这么做肯定有自身的道理，但是淘宝的电商平台和该电商公司的性质不同，而京东和亚马逊的模式与其更相像；但是这两家公司又没有淘宝做得规模大。

从这个问题来看，这个电商同仁思考问题的角度已经是非常深入了，对这个问题的方方面面都进行了考虑。当时，我并没有给他我的解决方案，因为我也不知道该怎么解决，于是我就只说了一句话：“先切出一部分流量来试试。”我这样做的目的是，我希望在还没有开局之前先能够有一些变化的东西呈现出来，然后用这个变化来进行破局。破局之后肯定就会有数据收集回来，有了数据之后，就有了一个可以参考的原材料。就像在格斗中那样，几下试探之后，高手一般就能够找到对方的攻防套路，就能够针对对方的套路进行有效的攻击。

然而，破局后问题也会随之而来：收集了数据之后又该怎么做？怎么样才算是好的，怎么样才算是不好的？这又是一个大问题。要解决这个问题，就要去衡量破局之后的每一个数据变化。

衡量每一个变化

在变化中，找到正确道路的方式是要能够去判断什么是对的，什么是错的。然后，把对的保持下去，把错的摒弃掉。如果无法有效地衡量这些变化，这样的变化就是无意义的。

我们可以再看一下之前提到的那个应该是突出“立即购买”，还是“加入购物车”的案例，如果数据出来了，最后测试的结论是突出“立即购买”的转化率要高于突出“加入购物车”的转化率，那这时，是不是就应该下结论说——这个电商网站就应该进行网页改版，做成像淘宝那样突出“立即购买”，或者至少也应该在“立即购买”和“加入购物车”上设置同样的比重。问题如果如此简单的话，那我想，思考就太过容易了。

在讨论这个指标之前，我们还有一个更加重要的目标是，让这个电商网站能够有更高的销售额。在通常情况下，电商有一个非常简单的公式：销售额=访问人数×转化率×客单价。所以，对于这次改动的衡

量，除了看转化率有没有下降之外，还需要关注另外两个指标。在这个案例中，会不会因为直接购买更加方便，而导致购买的笔数有所减少，使得单价有所降低？这个降低减弱了转化率提高的效果，甚至对销售额的增长起到了反向的作用。从人的角度来看，到底是哪些人对这样的改变比较敏感，哪些人不敏感，对通过广告投放引进来的人有没有什么改观等。所以，衡量就是要在目标确定的基础上，能够对一些变化做出有效的判断。就如在格斗中一样，我需要判断的是，对方的变化会对我的获胜有什么影响，我的出招对获胜有什么影响。如果无法衡量一次变动带来的各个方面的影响，那这样的变动就只能算是一次没有积累的变动，就有点自娱自乐的嫌疑了。

对于电商企业，衡量变动时需要有两个大的前提，一个前提是要保证数据是能够被有效地记录下来的，另一个前提是数据是能够被有效地解读的。不少产品经理在设计自己的产品时，可能会花费很多精力在产品本身的设计上，却没有花费精力去思考如何来衡量产品的成功与否。对于成功与否这个话题，通常只会在他们的产品文档中写上一句类似于“客户体验有所提升”的空话，这样的产品自然不会设计数据收集的点，自然也就没有可以衡量的行为可言。还有一类产品经理设计了数据收集的方案，但是这些数据并没有被实际运用或者只是被有限地使用，那么，这样的衡量也是缺失的，不利于进一步的改进。当然，就如格斗一样，一招之内决定胜负是不可能的，通常一次战斗会持续很多回合，所以我们需要对每一次变动都进行衡量，同时也要对每一次变动有所反馈，用这一数据作为下一次变动的基础。

对于整个格斗过程而言，反馈起到了链条的作用，只有通过反馈把变动串联在一起，才能让整个变动过程向着胜利的方向发展。

持续的反馈

在思考过程中，反馈是一个非常重要的环节。如果说格斗中的稳定来自于动态，那么思考中的稳定就来自于变化环境下的不断反馈。每一次反馈都会是迂回中一个新的起点，也会是一个新的衡量的开始。在持续的反馈过程中，整个思考过程将会变成一个有效的环路，这个有效的环路就如同自行车上的链条一样，能够将动力传导到车轮，从而带动车子向前行驶。

大数据环境瞬息万变，基本每一秒都会产生不同的规则。10年前，谁也不会预测到网络零售能够超过万亿规模；同样，10年后，我们也无法预知社会到底会变成什么样子。在面对飞速变化的环境时，对环境变化的适应尤为重要。就像在格斗中，面对对手不断变化的招式时，自己也需要不断地去调整自己的招式来应对。否则就会像程咬金的“三板斧”一样，虽然看起来威力无穷，但是一旦三招用完，就再也没有什么新东西了，只有催着他的宝马大肚子蝈蝈红逃跑了。我曾经和一些电商交流，在谈到运营数据的问题时，我指出很多时候，我们往往只对数据进行简单的衡量，并没有对数据进行持续的反馈。

所谓持续的反馈，就是在衡量的基础上，评价某一个变动到底是不是起到了很大的作用，并确定数据的变化确实是因为这次改变而出现的。现在，很多电商吸引用户的方式都是采用线上广告，这就是一个需要不断反馈的过程。比如今天我投放的媒体到底怎么样，可能一天的数据没法给出解释，需要通过很多天的数据来做决定。需要观察这个媒体的投放效果好不好，对哪些商品来说是好的，对哪些商品是不好的；效果好的商品是不是可以加大投入，效果不好的商品是不是需要更换媒体；对哪些用户是好的，对哪些用户又是不好的，等等。这都是需要解决的问题。对于一些新成立不久，还没有建立良好口碑的企业而言，这种不断的反馈将会在很短的时间内就建立起属于自己的经验体系。

寻找爆发点

像李小龙一样去做商业博弈，其中涉及很多方法，而这些方法的最终目的当然是要获取战斗的胜利，所有的迂回、衡量和反馈最终的目的就是为了寻找战机的出现。

我在上海认识一位研究国学的高人，我曾经向他请教过如何读《孙子兵法》这本书，他当时对《孙子兵法》做了一个非常精要的概括——战机。如果要读懂孙子兵法，就必须从中读懂如何去寻找战机，而《孙子兵法》中有很多章节与我们写到的迂回、衡量和反馈的方式非常接近。在《孙子兵法》中，有一句话叫作“多算胜，少算不胜，而况于无算乎”。用我们现在的话来讲，就是选择迂回前进，在运营数据中衡量每一个变动，并通过变动的结论来调整自己的下一个迂回动作。

对于“战机”，我想用一个简单的例子来说明。比如，有两辆坦克，

一个射程是3 000米，一个射程是300米（我不是军事专家，可能射程的说法有些不妥），如果射程300米的坦克想要获胜，那它应该如何去寻找战机呢？首先，这个射程为300米的坦克要知道对方的射程和己方的射程，如果它处在对方的射程之内、自己的射程之外，那只有当靶子的份儿；其次，要找到一个合适的方案，是迂回前进，还是做好伪装等待对方靠近；最后，如果在射程内，用什么方法击中敌方，这就是寻找战机的过程。当然，如果对手没有按照自己的假设来，则可以通过对手的一些情况不断地完善方案，最终寻找出一个正确的方案。

类似的情况在企业中也是屡见不鲜，我认为，未来的数据角色对企业而言就像决斗一样，如果两个人的水平差不多，多“算”的人就胜利了，而数据就是“算”的核心。但是，在寻找战机和“算”的过程中，还有一个需要非常注意的点，即分析方法一定要客观和理性，在经验不是特别丰富的情况下，千万不能盲目地用数据去观察问题，要保持相当清醒的头脑，不能让潜意识迷惑自己。

总的来说，战机就是基于不断的迂回、衡量和反馈，通过合理有效的逻辑思考方法来找到问题的关键点——问题的主干，因为抓住主干才能获得成功。

前面提到了很多点，都是在动态中去寻找战机的方法，而在这些动态中寻找到的战机，如果能够抽象出来，就是一件事情的主干。我认为的主干，就是一件事物最原始、最核心的规律。举一个简单的例子，我们说“打蛇打七寸”，“七寸”就是打蛇的一个主干，当然这个主干肯定是基于无数打蛇经验得出的。如果碰到一条蛇，我们没有打到七寸，那可能会受到它的攻击，甚至可能丧命；如果我们打到七寸，则可以轻松获得成功。我把这种抓住主干，然后通过主干解决一系列问题的情况，叫作“爆发”。但是有一点一定要明确，除非我们可以肯定自己的环境与别人是完全一样的，否则别人的爆发不一定就是我们自己的爆发。换句话说，别人的主干问题不是你的主干问题，所以简单套用解决方案的做法可能会害死自己。以前，传统企业一窝蜂地登录ERP系统，这些年IT公司又一窝蜂地搞云计算，就是简单套用别人的解决方案解决问题的失败案例。

在使用数据的过程中，其实也存在爆发。我之前写过一篇文章，提到了“数据扮演的三种角色”，分别是：第一个阶段为有多少数据就收

集多少数据，数据只从点上指导运营，并不指导战略和方向；第二个阶段为数据要精准化，数据之间开始有了关系，担当的角色从点到线或者面，就可以为战略作参考了；第三个阶段为做数据模型，分析外围数据开始变得异常重要，它可以为公司的下一步战略找到出路。在这三个阶段中，数据真正能够充当爆发角色的，一定是第三阶段。但是，如果没有前两个阶段收集和积累问题作为铺垫，那么，第三阶段的爆发点就不一定会出现，即便是真的出现了，也需要对其真实性进行考量。

爆发，是做每一件事情和解决每一个企业问题的人都需要考虑的，尤其是做数据的人，一定要时时刻刻想到数据能够为企业产生什么价值，能够用数据解决什么问题，是不是能够用数据来发现企业中的爆发点。如果做数据的人时刻都有这样的意识，那企业将会因为数据获得非常大的收益。

所以，我始终有一个看法，寻找爆发点的过程，其实就是自己和自己战斗的过程。在这之中需要不断忍受痛苦和折磨，不断尝试和精进，直到找到了自我突破点。

请记住，在很多时候，决斗对象不一定是他人，而是自己。

品觉的话

人在修行的路上，不要单打独斗

我相信一切善良的力量都会有善果，生命中那些有声和无声的人，都会以他们的方式存在、生长，在你和他人修行的道路上，看见彼此、发现彼此和祝福彼此。

结缘甘孜

七年前，我只身来到上海。有一天，忙累之余，我学佛的上师忽然问我："品觉，你要不要去甘孜看一下？"我当时没有想太多，就回答说："好啊，去就去吧。"是的，正是这一去完全改变了我后来的生活。

我的上师是雪谦寺的活佛，在我与他结缘时，甚至连甘孜的"孜"怎么念都不知道。上师很少和我说一些大道理，第一次去甘孜，他只是带着我，也不让我问行程。到了甘孜之后，上师让我去看看当地的孩子们。说实话，他们的生活状态给了我极大的冲击：五六十个孩子没有鞋子穿，也吃不饱，和老师们一起挤在破帐篷里。吃、住和教学都在 40平方米的破帐篷里，映入眼帘的只是他们破旧的行装和脸上的笑容……我当时什么都没有说，作为一个饱经沧桑的老男人，在那一刻，我竟然哭了。

我不骗你，现在回想起来，我都不知道自己为什么会流泪，也正是从那个时候起，我就下定决心：我要为这些孩子们做些什么，而且不能再等了。有些事情，可能是冥冥之中注定的，从2009年开始，我开始拉身边的人捐款，不够的就自己贴，以此保证每个月给100个学生每人100元钱的生活费。

你可能会问："为什么那个时候，我如此信任那些帮我们管理捐助金的人？"因为那个时候的管理者是五明佛学院的大湛布，他们负责帮助学校设计课程，管理孩子们的衣食住行，对于学佛的人而言，这一

点，非常可靠，也让我们这些捐助者没有后顾之忧。

自此，在工作之余，我的生命里忽然多了一股无法妥协的力量——如果月底不按时把钱汇过去，孩子们就会没饭吃。这种原动力，好像一个幸福的鞭子，一直在身后激励着我。

信任，电商人的温暖

甘孜受助的孩子中，有的是孤儿，有的是家境极其困难。2011年，我发现自己贴的钱越来越多，因为孩子们从100人增长到了200人，而且还在不断地增加。当时我日思夜想的就是，我必须要想办法让增加的孩子也能吃饱饭。

经常有人问我：为什么要自己掏钱，你也只是一个普通的打工者而已，这就是一个无底洞。我说不清楚，也答不上来，再说普通话也不好，也就无从为自己辩护。可是每一次去甘孜，看到孩子们的身体越来越好，老师们也很认真地在教学，再看到孩子们考试时自信的样子，我就知道，我必须坚持下去。这种力量，是亲眼所见带给我的，是外人无法体会的。

后来，我觉得不能太依靠身边人的捐助，我必须建立一个长期滚动的体系。因此，2011年，在家人及朋友的支持下，我开了淘宝店“桑珠贝玛”。这个店铺里主要是销售一些佛珠和茶叶，很多朋友也会将商品放到上面来义卖，所得款项全部捐给甘孜的孩子们。正是因为这个淘宝店，我开始养成早上四点半起床做“品觉晨读”的习惯，主要是为了店铺引流量，偶尔也做做客服，并利用空余时间去采购，当时我几乎是一个人把所有事情都搞定了。所以，从另一个侧面来看，我深深地了解到做一个淘宝卖家有多难。就这样不断坚持，之后店铺里每个月都有了七八千到一万元的收入，也算大大缓解了当时的危机。

但我明白，桑珠贝玛淘宝店能有如此的收入，大部分并不是因为我，而是归功于身边的好心人，而且我估计有一半的订单都是阿里巴巴的同事以及电商圈子里的朋友们下的。本来，在阿里巴巴，我作为一个高管，本人以及亲人都是不能开淘宝店的。在开店时，我对阿里巴巴廉政部说，如果我开店会给其他高管以借口，我宁可不开店。可是廉政部的同事对我说：“你继续做吧，只要每个月把收支公布出来，后面的事

情我们来处理。”听完他们的话，我觉得自己没有来错阿里巴巴。

就在前不久，忽然有人给我的支付宝捐了两万元，这两万元让我惊喜不已。后来我才知道，原来是我在支付宝时的老板和同事们捐的。他们现在常常开玩笑说：“以后同事之间打赌，输钱的一方都给车品觉助学用。”虽然只是一句玩笑话，但同事们却真的就这么做了，这让我很是感动。总之，阿里巴巴这家公司让我觉得自己没有来错。

还有一些电商圈的朋友，看都没看过甘孜的孩子，而且一次都没去过甘孜，却都在慷慨捐助。比如，唯品会的沈亚、亿邦动力网的郑敏、廖斌等好心人一直在暗中帮助我。有一次，唯品会的财务在做薪酬系统改动，到了月底捐助的钱还没有汇过来，沈亚对财务说，先把孩子们的钱给支出来，其他再说。他们对我没有任何怀疑，就这样选择相信我，这种信任是支撑我后来能够在助学路上一直坚持走下去的很重要的力量。

因为有一帮这样善良的人在帮助我，我更是要让每一元钱都对得起捐助的人。有时候，账目错了，少了点钱，我就自己补进去。佛教里面说：“别人捐的钱是公德，一毛钱都不能进自己的口袋。”这也就是很多时候，我们在讲善的用心。

做公益，第一重要的其实不是钱

在甘孜助学的过程中，我和我的一些朋友们开始明白，做公益，钱很重要，但真的不是最重要的，比钱更重要的还有两样东西。

一是要专业。记得2013年，一个朋友捐了上百个羽绒睡袋，那时恰逢天气忽然从十几度骤降到零下几度。说实话，物资从那么远的地方运到甘孜，其实很不容易。当我们发睡袋时，我的心温暖极了。从收到睡袋到发放，我们都有条不紊，而且大家配合得也很好。还有一次，我们给孩子们捐衣服，但是发到最后，恰好少了一件，有一个孩子没拿到，他哭得很厉害。当时，我们把自己身上的衣服脱给他，但他就是不肯要。可能有人会嗔怪，孩子不懂感恩。但事实并非如此，孩子是很单纯的，他只是想追求和普通人一样的东西。所以说，做公益真的不是一件容易的事情，其中需要计划得很细致，还要考虑到各个方面。因此，专业很重要，否则非常容易在过程中无意伤到孩子。

另一样重要的东西就是爱和用心。在甘孜助学的路上，这么多年，我有幸可以碰到许多有心的人。他们甚至从来不询问结果，只是选择无条件地信任。有很多人，捐钱捐东西，从来不会去寻求所谓的回馈和回报。在这些人心里，只有一个信念："你做了，我相信你，我相信只要对孩子们好，我们用心就好了。"这让我很感动。2010年，我们去甘孜的每个学校看孩子，很多学校连路都没有，中途迷路了，凌晨两点才到了一个小镇，安顿下来后我做了西红柿炒鸡蛋给同行的志愿者吃，他们说这是他们吃过最好吃的东西。这盘菜真的炒得那么好吗？不过是在那种情景下，大家都感觉有一种使命，必须找到那所学校，而围在一起吃一盘菜的温馨画面确实让人动容。

有很多人都在默默地做着很多事情，有的爱很直观，有的爱默默无闻。比如，我曾经和银泰网的Robin和今夜特价酒店的Mars说，我们自己是靠专业技能吃饭的，我的愿望就是，不仅是要让不能读书的孩子读书，还要让他们也能靠自己的专业技能谋生，这才是真正的好事情。随后，他们两个人都开始把出去讲课赚的钱捐了出来。

同行的很多人，总是令我觉得自己做得还不够。今年，我们再一次一起去甘孜，去之前我就在筹划写一本数据方面的书，筹划这本书的初衷是：一方面，利用自己这么多年做数据的经验和专业知识，把一些方法集合出来分享给大家；另一方面就是想筹集更多的钱来帮助更多的孩子吃饱饭，接受教育。而在这本书的筹备中，很多人都在出力。在去甘孜的路上，上师对我说："不忘初心，才会收获善果。"这句话，深深地植入我心，因为任何一个人善意的初心，一定会换来善果的。

我也改变了自己

我的一生并不像别人那样按部就班的学习、毕业、工作，一路走来我都是跌跌撞撞。我想，如果运气不好，现在的我很可能在路边帮人修车，在厨房给人炒菜，甚至加入了黑社会。我大学毕业时，连我老爸都哭了，他感觉这个孩子是被救回来的。我曾经有机会好好读书，但却没有，之后才浪子回头，我深深知道读书对于人有多重要，那是进入社会的门票。而现在，我看到这些孩子们连读书的机会都没有，顿觉世界的不公。这，大概就是我与这群孩子结缘的原因吧。

有人问我，是不是看到一个孩子在自己的帮助下成长，会很有成

就感。实际上，我们并没有特别关注哪个孩子，大家之间交流比较少，因为语言不通，只是每年会去看望他们一次。我最大的快乐就是，这些孩子能吃饱，而在吃饱的同时还能接受教育。

也正因为结了这样的缘，忙与累也变成了一种快乐。比如，这些年我每天坚持四点半起床，把每天重要的电商新闻摘出来发微博，因为我知道，每天只发一条淘宝店义卖的微博是不会有人看的，只有积累，大家才会看到你的用心。除此之外，我还养成了长期坚持早上阅读和打坐的习惯，这也让我对行业的了解更深，对于我的工作和生活来说都是好事情。

正因如此，上师对我说："你知道吗？2006年的品觉是一个很讨厌的人，现在却变化很大。"上师说的没错，在这个过程中，我也能感觉到自己心态的变化，我开始更少地去考虑自己，像上师说的，我开始学会把自己放在更低的位置，在这场修行中，我学会了放松，而世界似乎也更广阔了。

这是一个互联网的世界，每个人都不应该单打独斗。而我也还在修行的路上，一切还没有结束……

湛庐，与思想有关……

如何阅读商业图书

商业图书与其他类型的图书，由于阅读目的和方式的不同，因此有其特定的阅读原则和阅读方法，先从一本书开始尝试，再熟练应用。

二八原则

对商业图书来说，80%的精华价值可能仅占20%的页码。要根据自己的阅读能力，进行阅读时间的分配。

阅读原则**2** 集中优势精力原则

在一个特定的时间段内，集中突破20%的精华内容。也可以在一个时间段内，集中攻克一个主题的阅读。

阅读原则3 递进原则

高效率的阅读并不一定要按照页码顺序展开，可以挑选自己感兴趣的部分阅读，再从兴趣点扩展到其他部分。阅读商业图书切忌贪多，从一个小主题开始，先培养自己的阅读能力，了解文字风格、观点阐述以及案例描述的方法，目的在于对方法的掌握，这才是最重要的。

阅读原则**4** 好为人师原则

在朋友圈中主导、控制话题，引导话题向自己设计的方向去发展，可以让读书收获更加扎实、实用、有效。

阅读方法与阅读习惯的养成

（1）回想。阅读商业图书常常不会一口气读完，第二次拿起书时，至少用15分钟回想上次阅读的内容，不要翻看，实在想不起来再翻看。严格训练自己，一定要回想，坚持50次，会逐渐养成习惯。

（2）做笔记。不要试图让笔记具有很强的逻辑性和系统性，不需要有深刻的见解和思想，只要是文字，就是对大脑的锻炼。在空白处多写多画，随笔、符号、涂色、书签、便签、折页，甚至拆书都可以。

（3）读后感和PPT。坚持写读后感可以大幅度提高阅读能力，做PPT可以提高逻辑分析能力。从写读后感开始，写上5篇以后，再尝试做PPT。连续做上5个PPT，再重复写三次读后感。如此坚持，阅读能力将会大幅度提高。

（4）思想的超越。要养成上述阅读习惯，通常需要6个月的严格训练，至少完成4本书的阅读。你会慢慢发现，自己的思想开始跳脱出来，开始有了超越作者的感觉。比拟作者、超越作者、试图凌驾于作者之上思考问题，是阅读能力提高的必然结果。

好的方法其实很简单，难就难在执行。需要毅力、执著、长期的坚持，从而养成习惯。用心学习，就会得到心的改变、思想的改变。阅读，与思想有关。

［特别感谢：营销及销售行为专家 孙路弘 智慧支持！］

我们出版的所有图书，封底和前勒口都有“湛庐文化”的标志

并归于两个品牌

- 找**“**小红帽**”**

为了便于读者在浩如烟海的书架陈列中清楚地找到湛庐，我们在每本图书的封面左上角，以及书脊上部47mm处，以红色作为标记——称之为**“**小红帽**”**。同时，封面左上角标记**“**湛庐文化**Slogan”**，书脊上标记**“**湛庐文化**Logo”**，且下方标注图书所属品牌。

湛庐文化主力打造两个品牌：财富汇，致力于为商界人士提供国内外优秀的经济管理类图书；心视界，旨在通过心理学大师、心灵导师的专业指导为读者提供改善生活和心境的通路。



- 阅读的最大成本

读者在选购图书的时候，往往把成本支出的焦点放在书价上，其实不然。

时间才是读者付出的最大阅读成本。

阅读的时间成本=选择花费的时间+阅读花费的时间+误读浪费的时间

湛庐希望成为一个“与思想有关”的组织，成为中国与世界思想交汇的聚集地。通过我们的工作和努力，潜移默化地改变中国人、商业组织的思维方式，与世界先进的理念接轨，帮助国内的企业和经理人，融入世界，这是我们的使命和价值。

我们知道，这项工作就像跑马拉松，是极其漫长和艰苦的。但是我们有决心和毅力去不断推动，在朝着我们目标前进的道路上，所有人都是同行者和推动者。希望更多的专家、学者、读者一起来加入我们的队伍，在当下改变未来。

湛庐文化2008-2012年获奖书目

- 《正能量》

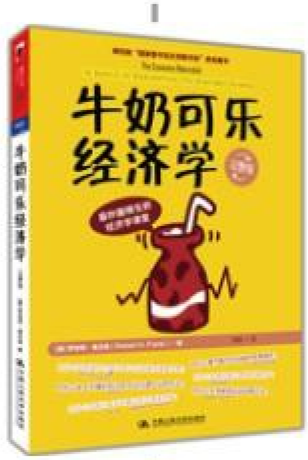


《新智囊》2012年经管类十大图书，京东2012好书榜年度新书。

35年职业经理人养成心得，写给有追求的职场人。

聆听总裁的职场故事，发掘自己与生俱来的正能量。

- 《牛奶可乐经济学》



国家图书馆“第四届文津奖”十本获奖图书之一，唯一获奖的商业类图书。

搜狐、《第一财经日报》2008年十本最佳商业图书。

用经济学的眼光看待生活和工作，体验作为“经济学家”的美妙之处。

- 《清单革命》



《中国图书商报》商业类十大好书。

全球思想家正在读的20本书之一。

一场应对复杂世界的观念变革，一部捍卫安全与正确的实践宣言。

- 《大而不倒》



《金融时报》·高盛2010年度最佳商业图书入选作品。

美国《外交政策》杂志评选的全球思想家正在阅读的20本书之一。

蓝狮子·新浪2010年度十大最佳商业图书，《智囊悦读》2010年度十大最具价值经管图书。

一部金融界的《2012》，一部丹·布朗式的鸿篇巨制。

- 《金融之王》



《金融时报》·高盛2010年度最佳商业图书。

蓝狮子2011年度十大最佳商业图书，《第一财经日报》2011年度十大金融投资书籍。

一部优美的人物传记，一部独特视角的经济金融史。

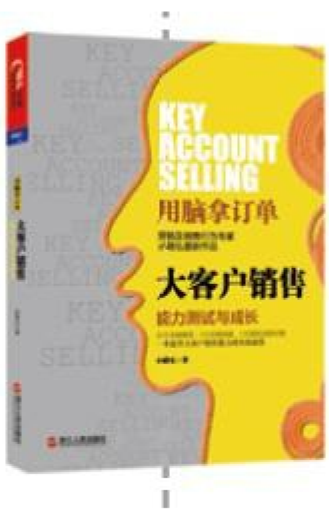
- 《快乐竞争力》



蓝狮子2012年度十大最佳商业图书。

赢得优势的7个积极心理学法则，全美10大幸福企业“幸福感”培训专用书。

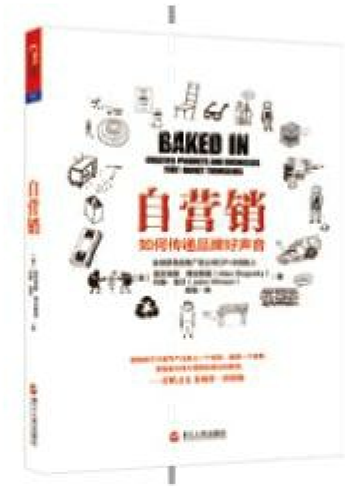
- 《大客户销售》



蓝狮子·新营销2012最佳营销商业图书。

著名营销及销售行为专家孙路弘最新作品，一本提升大客户销售能力的实战秘笈。

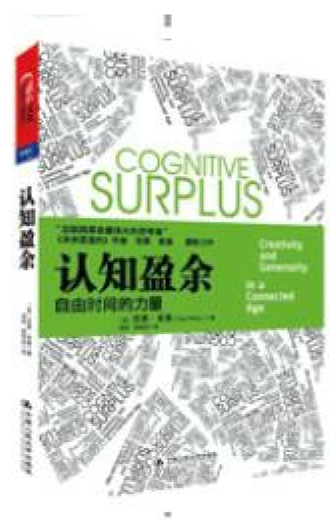
- 《自营销》



百道网2013年度潜力新书。

全球最具创意广告公司CP+B掌门人的洞见之作，让好产品和好营销同唱一首歌。

- 《认知盈余》

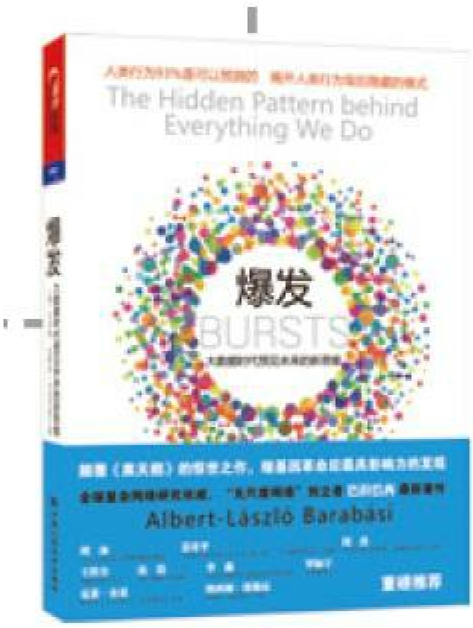


2011年度和讯华文财经图书大奖。

看“互联网革命最伟大的思考者”克莱·舍基如何开启无组织的时间力量。

看自由时间如何成就“有闲”世界，如何引领“有闲”经济与“有闲”商业的未来。

- 《爆发》



百道网2013年度潜力新书。

大数据时代预见未来的新思维，颠覆《黑天鹅》的惊世之作，揭开人类行为背后隐藏的模式。

- 《微力无边》



2011年度和讯华文财经图书大奖“最佳装帧设计奖”。

中国最早的社会化媒体营销研究者杜子建首部作品，一部微博前传，半部营销后传。

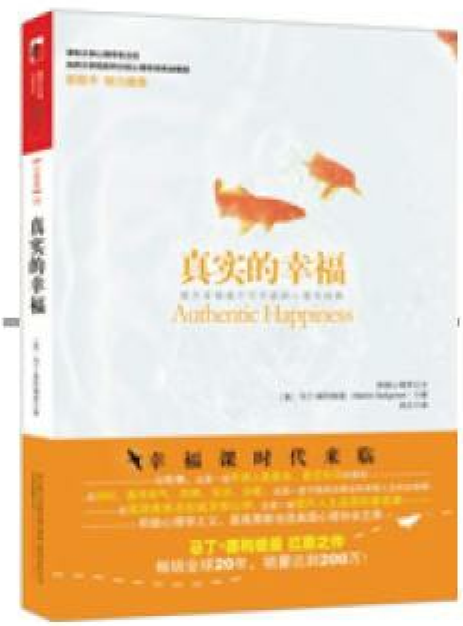
- 《神话的力量》



《心理月刊》2011年度最佳图书奖。

在诸神与英雄的世界中发现自我，当代神话学大师约瑟夫·坎贝尔毕生精髓之作。

- 《真实的幸福》



《职场》2010年度最具阅读价值的10本职场书籍。

积极心理学之父马丁·塞利格曼扛鼎之作。

哈佛最吸引人、最受欢迎的幸福课。

延伸阅读

《大数据时代：生活、工作与思维的大变革》

◎ “大数据时代的预言家”维克托·迈尔–舍恩伯格力作。

◎ 迄今为止全世界最好的一本大数据专著。

◎ 国外大数据系统研究的先河之作，一场生活、工作与思维的大变革。



扫码直达本书购买链接

《删除：大数据取舍之道》

◎ “大数据时代的预言家”维克托·迈尔–舍恩伯格关于大数据的最经典著作。

◎ 本书洞见了“被遗忘的权利”，探索了大数据时代人类该如何面对海量信息人类该如何取舍，以构建一个积极而安全的未来。

◎ “马歇尔·麦克卢汉奖”获奖图书，《华尔街日报》《连线》《科学》杂志书评推荐。



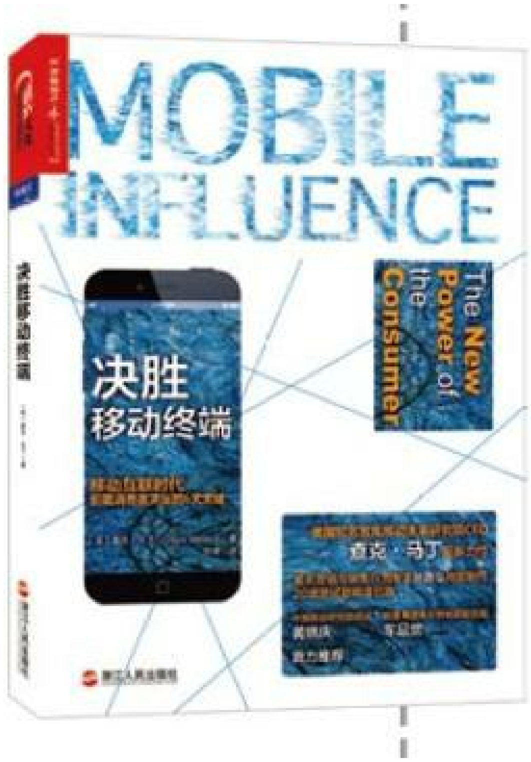
扫码直达本书购买链接

《决胜移动终端：移动互联时代影响消费者决策的**6**大关键》

◎ 美国知名智库移动未来研究院CEO查克·马丁最新力作。

◎ 第一本真正阐释O2O概念的移动互联必读之作，是传统企业拥抱移动互联网的经验大典！

◎ 中国移动研究院院长黄晓庆，阿里巴巴副总裁车品觉鼎力推荐！



扫码直达本书购买链接

《大数据云图：如何在大数据时代寻找下一个大机遇》

◎ 作者大卫·芬雷布被誉为“大数据商业应用的引路人”。

◎ 讲述亚马逊、Google、IBM、Facebook、LinkedIn…… 超过一百家大数据公司的商业法则深度解密。大数据云图清晰勾勒出大数据行业的企业分布，让你能够轻易发现大数据行业的下一个大机遇究竟在何处。


BIG DATA
DEMYSTIFIED
HOW BIG DATA IS CHANGING
THE WAY WE LIVE, LOVE AND LEARN
大数据云图
如何在大数据时代寻找
下一个大机遇

扫码直达本书购买链接



你需要知道的桑珠助学

十年前，拥珠扎西仁波切在如意宝法王面前许下愿望，希望能让更多的儿童获得受教育的机会，法王当下为学校赐名"桑珠"。"桑珠"在藏语中是"心想事成"的意思。2010 年开始，拥珠扎西仁波切的弟子车品觉受命为第一批分布在五个地方的 296 名学生筹集经费与师资。现在，在藏区负责管理和分配物资的负责人以五明佛学院堪布为主。

目前，已开办的学校分布在甘孜州各处，有些在偏远的达通玛大草原上，分别是阿卡寺学校、苏坡尼斯学校、雪域利众学校、阿色寺学校、东托寺学校、阿德学校、折查利众学校、东果里乌寺学校，每个学校大约有几十名学生，有一位老师与孩子们同吃同住，有一位负责做饭等事务的喇嘛，所有孩子住在既是教室又是卧室的屋子里，老师和另一个工作人员则住在堆放着食物和炉灶的厨房；有的学校校舍就是寺庙，有的校舍还在修建中，孩子们和老师都暂住在帐篷里。

车品觉以及其余 7 位参与桑珠助学计划并资助藏区儿童的朋友已去往甘孜州的各个学校，跟随堪布在学校发放物资、经费，参与了堪布与各学校负责人商谈教学管理事宜及经费物资领取签字盖章等活动。每位进入学校的孩子，其家庭情况都实时记录在册，虽然还在摸索阶段，但毫无疑问的是，每位参与者都怀着最无私的爱心和责任心，一切都在走向更正规、更系统、更长久的发展道路上，我们相信一定会帮助到更多地方的更多孩子，让他们在更好的环境里接受更好的教育。在这些人眼中，每个孩子都是天使，他们的纯真和求知欲值得每个人伸手相助。事实上，你帮助的不只是一两个或几个可爱的孩子，也是在展现自己内心最深处的爱心和感动。

如果你能感受到这份不同于你身边所见且真实的另一个世界，愿意加入到帮助这些孩子们的行列中，可以联系拥珠扎西仁波切的弟子车品觉，我们会尽可能详尽地让你了解桑珠行动的情况。

目前，每个孩子每年的资助费用为 1200 元人民币，这笔费用已分摊至教学场地、教具、师资（师资来自中国最大的佛学院五明佛学院）、食品衣物、日常用品等方面。资助金用途可靠，如果你有更多的爱心或疑问，可以到当地探望你资助的小朋友们。

如果我们贡献微薄之力，携手共同努力，聚集的力量一定能为孩子们的将来做出更多更大的改变。



桑珠助学编年史

2000 年

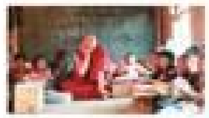


拥珠扎西仁波切在恩师如意宝法王面前许下心愿，终生致力于改善山区失学儿童的教育，如意宝法王赐名“桑珠”，意思是“心想事成”。

2006 年

筹建甘孜利民福利院，但因种种情况，此项目搁浅。

2007 年



第一所桑珠学校成立，学生约 50 人。

2008 年

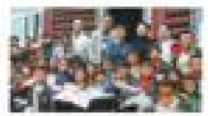


桑珠学校陆续筹建以及完善中，学生增至约 150 人。

2009 年

车品觉正式开始为桑珠学校学生们募集生活费用，学生增至约 200 人。

2010 年



7 月，“桑珠行”走访 8 所学校，学生总数约 300 人；

11 月，为义卖筹建淘宝“桑珠贝玛”小店。

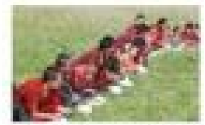
2011 年



5 月，在德格县阿须草原，全体同学第一次参加桑珠会考，学生总数约 400 人；

10 月，淘宝“桑珠贝玛”小店正式成立。

2012 年



4 月，10 所学校的学生们集中在色达五名佛学院的草地上进行第二次桑珠会考，学生总数约 500 人；

6 月，第一只基金“唯品会爱心基金”成立，其每月按时捐助桑珠助学；

9 月，“桑珠行”探望阿坝自治州山区内新建学校。

2013 年



8 月，第三次桑珠会考，“桑珠行”走访 9 所学校。各所学校分别进行考试，学生总数约 610 人。